



# UNIVERSITÀ DEGLI STUDI DI PALERMO

Dottorato in Energia indirizzo Energetica  
Dipartimento di Energia, Ingegneria dell'Informazione e Modelli Matematici  
Settore Scientifico Disciplinare ING-IND/11

## ANALISI E MODELLI PREDITTIVI DI SISTEMI ENERGETICI A SCALA REGIONALE, PROVINCIALE E LOCALE: STUDI SPERIMENTALI, MODELLAZIONE E ANALISI PARAMETRICA.

IL DOTTORE  
**Ing. Edoardo Moreci**

IL COORDINATORE  
**Ch.mo Prof. Ing. Aldo Orioli**

IL TUTOR  
**Prof. Ing. Valerio Lo Brano**

CICLO XXVI  
ANNO CONSEGUIMENTO TITOLO 2016



# Sommario

|   |           |
|---|-----------|
| <b>CAPITOLO 1 Premessa .....</b>  | <b>1</b>  |
| <b>CAPITOLO 2 Le fasi della ricerca .....</b>   | <b>3</b>  |
| 2.1 Step 1 – Ricerca bibliografica e normativa di settore .....                               | 3         |
| 2.2 Step 2 – Metodi di analisi e predizione dei dati storici .....                            | 5         |
| 2.3 Step 3 – Reperimento ed organizzazione del dato energetico .....                          | 6         |
| 2.4 Step 4 – Implementazione dei casi studio .....  | 6         |
| <b>CAPITOLO 3 Ricerche bibliografiche .....</b>   | <b>8</b>  |
| <b>CAPITOLO 4 Le statistiche energetiche, evoluzione della normativa di riferimento .....</b> | <b>14</b> |
| 4.1 Contesto mondiale .....   | 14        |
| 4.2 L'Europa e le statistiche sull'energia .....  | 15        |
| 4.3 Le statistiche dell'energia in Italia .....   | 17        |
| <b>CAPITOLO 5 I dati statistici sull'energia.....</b>   | <b>20</b> |
| 5.1 Fonte primaria e secondari di energia .....   | 20        |
| 5.2 Fonti rinnovabili e non rinnovabili .....   | 20        |
| 5.3 Le unità di misura dell'energia .....   | 21        |
| 5.4 I dati economici correlati all'energia .....  | 22        |
| 5.5 I dati climatici .....  | 24        |
| 5.6 I dati demografici.....   | 25        |
| 5.7 Gli indicatori energetici ed ambientali .....   | 25        |
| 5.7.1 Fattore di intensità energetica .....   | 26        |
| 5.7.2 Gli indicatori ambientali .....   | 27        |
| <b>CAPITOLO 6 Analisi di dati temporali.....</b>  | <b>31</b> |
| 6.1 Le serie storica e dati cross-section .....   | 32        |
| 6.2 Le fasi di lavoro della predizione .....  | 33        |
| 6.3 Analisi grafiche preliminari .....  | 34        |
| 6.4 Analisi descrittive dei dati .....  | 35        |
| 6.5 Trasformazioni e aggiustamenti .....  | 36        |
| 6.6 Metodi di previsione.....   | 36        |
| 6.6.1 Metodi quantitativi e qualitativi .....   | 37        |
| 6.6.2 Approccio classico e moderno .....  | 38        |
| 6.6.3 Metodi informali - qualitativi.....   | 39        |
| 6.6.4 Metodo di scomposizione .....   | 40        |
| 6.6.5 Metodi di lisciamiento esponenziali.....  | 42        |
| 6.6.6 Metodo di regressione.....  | 46        |
| 6.7 I modelli ARIMA .....   | 47        |
| 6.8 Group method of data handling - GMDH .....  | 51        |
| 6.8.1 Algoritmo combinatorio (COMBI) .....  | 52        |
| 6.8.2 Algoritmi multistrato.....  | 53        |
| 6.8.3 Algoritmo Fuzzy .....   | 54        |

|  |           |
|--|-----------|
| 6.8.4 Algoritmo non parametrico .....                                      | 55        |
| 6.8.5 Algoritmi semiparametrici .....                                      | 57        |
| 6.9 Misure di bontà di adattamento e di accuratezza della previsione ..... | 57        |
| <b>CAPITOLO 7 Le reti neurali .....</b>                                    | <b>59</b> |
| 7.1 Introduzione alle reti neurali .....                                   | 59        |
| 7.2 Il neurone artificiale .....   | 61        |
| 7.3 Funzioni di attivazione .....  | 62        |
| 7.4 Tipologie di reti neurali .....  | 65        |
| 7.5 Modalità di utilizzo .....   | 65        |
| 7.6 Modalità di apprendimento .....  | 66        |
| 7.7 Architettura dei collegamenti .....                                    | 69        |
| 7.8 Algoritmi di apprendimento .....                                       | 69        |
| 7.8.1 Regola di Hebb .....   | 69        |
| 7.8.2 Regola postsinaptica .....   | 70        |
| 7.8.3 Regola presinaptica .....  | 70        |
| 7.8.4 Regola della covarianza .....  | 71        |
| 7.8.5 Regola di apprendimento del percettone semplice .....                | 71        |
| 7.8.6 Regola delta .....   | 72        |
| 7.8.7 Algoritmo di back-propagation e reti neurali multistrato .....       | 76        |
| 7.9 Problemi nella generalizzazione .....                                  | 79        |
| 7.10 Il modello GMDH e le reti neurali .....                               | 81        |
| 7.10.1 GMDH type Neural Networks .....                                     | 82        |
| <b>CAPITOLO 8 I Bilanci energetici locali .....</b>                        | <b>84</b> |
| 8.1 Il Bilancio Energetico locale: aspetti metodologici .....              | 85        |
| 8.2 La struttura generale del B.E.L. ....                                  | 87        |
| 8.2.1 Le righe .....   | 88        |
| 8.2.2 Le colonne .....   | 91        |
| <b>CAPITOLO 9 I software utilizzati .....</b>                              | <b>92</b> |
| 9.1 Pacchetto IBM SPSS .....   | 92        |
| 9.1.1 Preparazione dei dati .....  | 93        |
| 9.1.2 Trasformazioni di dati .....   | 93        |
| 9.1.3 Serie storiche .....   | 94        |
| 9.1.4 L'output .....   | 96        |
| 9.1.5 I grafici .....  | 96        |
| 9.1.6 Le statistiche descrittive dei dati .....                            | 97        |
| 9.2 GMDH Shell (Forecasting Software for Business) .....                   | 98        |
| 9.2.1 Il DataSet .....   | 98        |
| 9.2.2 Le analisi sul DataSet .....   | 99        |
| 9.2.3 Pre-processazione del DataSet .....                                  | 103       |
| 9.2.4 Il Set del risolutore .....  | 107       |
| 9.2.5 Il post-processor .....  | 108       |
| 9.2.6 L'analisi dei risultati .....  | 109       |



|   |            |
|---|------------|
| <b>CAPITOLO 10 Caso Studio Consumi energetici.....</b>                        | <b>111</b> |
| 10.1 Scala regionale.....   | 111        |
| 10.1.1 Il DataSet.....  | 111        |
| 10.1.2 Analisi del DataSet .....  | 113        |
| 10.1.3 Modello di previsione Neural Network “Consumi EE regionali” .....      | 118        |
| 10.1.4 Modello di previsione serie storiche “Consumi EE regionali” .....      | 124        |
| 10.2 Scala provinciale.....   | 129        |
| 10.2.1 Il DataSet.....  | 129        |
| 10.2.2 Analisi del DataSet .....  | 134        |
| 10.2.3 Modello di previsione Neural Network “Consumi EE provinciali” .....    | 137        |
| 10.3 Scala comunale .....   | 143        |
| 10.3.1 Il DataSet.....  | 144        |
| 10.3.2 Analisi del DataSet .....  | 145        |
| 10.3.3 Modello di previsione Neural Network “Consumi EE e gas comunali” ..... | 148        |
| 10.4 Conclusioni.....   | 154        |
| <b>CAPITOLO 11 Caso studio Produzione di energia elettrica da FER.....</b>    | <b>156</b> |
| 11.1 Il contesto della regione Sicilia .....                                  | 156        |
| 11.2 I dati energetici .....  | 167        |
| 11.3 I dati meteo climatici .....   | 168        |
| 11.4 Il DataSet.....  | 169        |
| 11.5 Analisi del DataSet .....  | 170        |
| 11.6 Modello di Previsione Neural Network “Produzione EE da FER” .....        | 175        |
| 11.7 Modello di Previsione serie storiche “Produzione EE da FER” .....        | 181        |
| 11.8 Conclusioni.....   | 184        |
| <b>CAPITOLO 12 Caso studio Prezzo dell’energia elettrica.....</b>             | <b>186</b> |
| 12.1 I problemi indotti dalle rinnovabili al mercato .....                    | 187        |
| 12.2 La base dati.....  | 188        |
| 12.3 L’analisi dati.....  | 189        |
| 12.4 Modello di Previsione Neural Network “Prezzo zonale” .....               | 202        |
| 12.5 Modello di Previsione serie storiche “Prezzo zonale” .....               | 206        |
| 12.6 Conclusioni.....   | 208        |
| <b>CAPITOLO 13 Caso studio Edificio uso ufficio.....</b>                      | <b>209</b> |
| 13.1 La bibliografia sugli edifici uso ufficio .....                          | 210        |
| 13.2 Gli stati oggetto dello studio .....                                     | 211        |
| 13.3 Il modello di simulazione.....   | 216        |
| 13.4 Le simulazioni e la validazione.....                                     | 222        |
| 13.5 Modello di regressione $H_d$ -HDD .....                                  | 223        |
| <b>CAPITOLO 14 Conclusioni.....</b>   | <b>227</b> |
| <b>Bibliografia .....</b>   | <b>231</b> |

## *Indice delle tabelle*

Tabella 5.1 - Fattore di conversione in tep dei combustibili solidi - Fonte: E.N.E.A

Tabella 5.2 - Fattore di conversione in tep dei combustibili liquidi - Fonte: E.N.E.A.

Tabella 5.3 - Fattore di conversione in tep dei combustibili gassosi - Fonte: E.N.E.A.

Tabella 9.1 - Funzioni elementari - GMDH Shell.

Tabella 9.2 - Funzioni per serie temporali - GMDH Shell.

Tabella 9.3 - Funzioni per set di dati temporali - GMDH Shell.

Tabella 9.4 - Funzioni calendario - GMDH Shell.

Tabella 9.5 - Funzioni per calibrazione - GMDH Shell.

Tabella 9.6 - Variabili speciali - GMDH Shell.

Tabella 9.7 - Modalità di convalida del modello - GMDH Shell.

Tabella 10.1 – Valori climatici, economici ed ambientali DataSet Sicilia.

Tabella 10.2 – Consumi energetici DataSet Sicilia.

Tabella 10.3 – Indicatori energetici DataSet Sicilia.

Tabella 10.4 – Indicatori ambientali DataSet Sicilia.

Tabella 10.5 – Statistiche descrittive valori climatici, economici ed ambientali DataSet Sicilia.

Tabella 10.6 – Statistiche descrittive consumi energetici DataSet Sicilia.

Tabella 10.7 – Statistiche indicatori energetici DataSet Sicilia.

Tabella 10.8 – Statistiche indicatori ambientali DataSet Sicilia.

Tabella 10.9 – Settings solver – Modello di previsione “Consumi EE regionali”.

Tabella 10.10 – Accuracy ktep residenziale Modello di previsione “Consumi EE regionali” con NN.

Tabella 10.11 – Accuracy ktep servizi non vendibili Modello di previsione “Consumi EE regionali” con NN.

Tabella 10.12 – Accuracy ktep agricoltura Modello di previsione “Consumi EE regionali” con NN.

Tabella 10.13 – Accuracy ktep industria Modello di previsione “Consumi EE regionali” con NN.

Tabella 10.14 – Accuracy ktep terziario Modello di previsione “Consumi EE regionali” con NN.

Tabella 10.15 – Accuracy ktep attività produttive Modello di previsione “Consumi EE regionali” con NN.

Tabella 10.16 – Settings solver – Modello di previsione “Consumi EE regionali”.

Tabella 10.17 – Accuracy ktep residenziale– Modello di previsione “Consumi EE regionali”.

Tabella 10.18 – Accuracy ktep servizi non vendibili– Modello di previsione “Consumi EE regionali”.

Tabella 10.19 – Accuracy ktep agricoltura– Modello di previsione “Consumi EE regionali”.

Tabella 10.20 – Accuracy ktep industria– Modello di previsione “Consumi EE regionali”.

Tabella 10.21 – Accuracy ktep terziario– Modello di previsione “Consumi EE regionali”.

Tabella 10.22 – Accuracy ktep attività produttive– Modello di previsione “Consumi EE regionali”.

Tabella 10.23 – Valori climatici, economici ed ambientali DataSet Provincie Sicilia.

Tabella 10.24 – Consumi energetici DataSet Provincie Sicilia

Tabella 10.25 – Indicatori energetici DataSet Provincie Sicilia.

Tabella 10.26 – Indicatori ambientali DataSet Provincie Sicilia.

Tabella 10.27 – Statistiche descrittive valori climatici, economici ed ambientali DataSet Provincie Sicilia.

Tabella 10.28 – Statistiche descrittive consumi energetici DataSet Provincie Sicilia.

Tabella 10.29 – Statistiche indicatori energetici DataSet Provincie Sicilia

Tabella 10.30 – Statistiche indicatori ambientali DataSet Provincie Sicilia.

Tab. 10.31 – Settings solver – Modello di previsione “Consumi EE provinciali”.

Tabella 10.32 – Accuracy ktep residenziale Modello di previsione “Consumi EE provinciali” con NN.

Tabella 10.33 – Accuracy ktep agricoltura Modello di previsione “Consumi EE provinciali” con NN.

Tabella 10.34 – Accuracy ktep industriaModello di previsione “Consumi EE provinciali” con NN.

Tabella 10.35 – Accuracy ktep terziario Modello di previsione “Consumi EE provinciali” con NN.

Tabella 10.36 – Dati input DataSet Comuni Sicilia.

Tabella 10.37 – Dati output DataSet Comuni Sicilia.

Tabella 10.38 – Statistiche descrittive valori input DataSet Comuni Sicilia.

Tabella 10.39 – Statistiche descrittive output DataSet Comuni Sicilia.

Tabella 10.40 – Settings solver – Modello di previsione “Consumi EE e gas comunali”.

Tabella 10.41 – Accuracy tep energia elettrica Modello di previsione “Consumi EE e gas comunali” con NN.

Tabella 10.42 – Accuracy tep gas naturale Modello di previsione “Consumi EE e gas comunali” con NN.

Tabella 11.1 - Potenza installata Sicilia – Italia anno 2014.

Tabella 11.2 - Potenza impianti FER – Sicilia [MW].

Tabella 11.3 - Produzione impianti FER – Sicilia [GWh].

Tabella 11.4 - DataSet energetico.

Tabella 11.5 - DataSet climatico.

Tabella 11.6 - DataSet utilizzato nei modelli di previsione.

Tabella 11.7 - Statistiche e caratteristiche DataSet parte energetica

Tabella 11.7 - Sstatistiche e caratteristiche DataSet parte climatica.

Tabella 11.8 – Settings solver – Modello di previsione “Produzione EE da FER”.

Tabella 11.9 – Accuracy PV tep– Modello di previsione “Produzione EE da FER” con NN.

Tabella 11.10 – Accuracy Eolico tep– Modello di previsione “Produzione EE da FER” con NN.

Tabella 11.11 – Accuracy Idro fl tep– Modello di previsione “Produzione EE da FER” con NN.

Tabella 11.12 – Settings solver – Modello di previsione “Produzione EE da FER”.

Tabella 11.13 – Accuracy PV tep– Modello di previsione “Produzione EE da FER”.

Tabella 11.14 – Accuracy Eolico tep– Modello di previsione “Produzione EE da FER”.

Tabella 11.15 – Accuracy Idro fl tep– Modello di previsione “Produzione EE da FER”

Tabella 12.1 - DataSet.

Tabella 12.2 - Statistiche descrittive DataSet.

Tabella 12.3 - Potenza installata in MW per singola fonte dal 2011 al 2014.

Tabella 12.3 - Statistiche descrittive DataSet anno 2011.

Tabella 12.4 - Statistiche descrittive DataSet anno 2012.

Tabella 12.5 - Statistiche descrittive DataSet anno 2013.

Tabella 12.6 - Statistiche descrittive DataSet anno 2014.

Tabella 12.7 – Settings solver – Modello di previsione “Prezzo Zonale”.

Tabella 12.8 – Accuracy – Modello di previsione “Offerta di Vendita” con NN.

Tabella 12.9 – Settings solver – Modello di previsione “Prezzo Zonale”.

Tabella 12.10 – Accuracy – Modello di previsione “Prezzo Zonale”.

Tabella 12.11 – Confronto modelli di previsione “Prezzo Zonale”.

Tabella 13.1 - HDD, CDD e radiazione solare degli Stati europei.

Tabella 13.2 - Valori di HDD, CDD e radiazione solare per gli Stati oggetto dello studio.

Tabella 13.3 - HDD range.

Tabella 13.4 - Dati empirici sull'energia termica richiesta dagli edifici uso ufficio negli Stati oggetto dello studio

Tabella 13.5 - Città prese ad esame nello studio.

Tabella 13.6 - Somma dei contributi gratuiti di occupanti e apparecchiature.

Tabella 13.7 - Contributi gratuiti di occupanti e apparecchiature.

Tabella 13.8 - Valori medi di trasmittanza per pareti pavimenti e tetti per valore di HDD nei singoli stati.

Tabella 13.9 - Valori U-Value delle finestre.

Tabella 13.10 - Valori di HDD massimi, minimi e medi e periodi di riscaldamento delle città oggetto dello studio.

Tabella 13.11 - Valori termofisici utilizzati nello sviluppo del modello su TRNSYS.

Tabella 13.12 - Confronto tra dati simulate e dati empirici.

Tabella 14.1 – Indici di bontà dei modelli serie storiche per i casi del paragrafo 10.1 e capitoli 11 e 12.

Tabella 14.2 – Indici di bontà del model fit modelli NN per i casi dei paragrafi 10.1 e 10.2.

Tabella 14.3 – Indici di bontà del model fit dei modelli NN per i casi del paragrafo 10.3 e capitoli 11, 12.

Tabella 14.4 – Indici di bontà della prediction dei modelli NN per i casi dei paragrafi 10.1 e 10.2.

Tabella 14.5 – Indici di bontà della prediction dei modelli NN per i casi del paragrafo 10.3 e capitoli 11 e 12.

## *Indice delle figure*

Figura 6.1 - Time plot di serie storiche di dati.

Figura 6.2 - Esempio di serie trimestrale a componenti additivi.

Figura 6.3 - Esempio di una serie trimestrale a componenti moltiplicativi.

Figura 6.4 - GMDH Algoritmo multistrato.

Figura 7.1 - Rappresentazione del neurone e dello strato di neuroni.

Figura 7.2 - Rappresentazione del neurone e dello strato di neuroni rappresentativi delle reti neurali artificiali.

Figura 7.3 - Neurone artificiale.

Figura 7.4 - Neurone Artificiale più Unità di Bias (cerchio nero).

Figura 7.5 - Funzione di attivazione a gradino.

Figura 7.6 - Funzione bipolare.

Figura 7.7 - Funzione lineare continua.

Figura 7.8 - Funzione sigmoidea.

Figura 7.9 - Tangente iperbolica.

Figura 7.10 - Schema semplificato di apprendimento supervisionato.

Figura 7.11 - Schema apprendimento supervisionato.

Figura 7.12 - Schema semplificato di apprendimento non supervisionato.

Figura 7.13 - Schema di apprendimento non supervisionato.

Figura 7.14 - Esempio di architettura dei collegamenti di una rete neurale.

Figura 7.15 - Tasso di apprendimento troppo grande.

Figura 7.16 - Tasso di apprendimento troppo piccolo

Figura 7.17 - Superficie della funzione di errore EW per una rete multistrato con unità continue non-lineari.

Figura 7.18 - Rete multistrato feed-forward con unità non lineari.

Figura 7.19 - Apprendimento della funzione che descrive i dati di addestramento (cerchietti neri). L'errore è lo scarto quadratico medio tra un dato e il corrispondente punto della funzione. a) Un numero troppo grande di parametri non garantisce una buona generalizzazione del nuovo pattern (cerchietto bianco); b) un numero adeguato di parametri permette una migliore generalizzazione (lo scarto quadratico medio è inferiore).

Figura 7.20 - Andamento dell'errore sui pattern di addestramento e di test (più i pattern di validazione); il livello ottimale di prestazioni si ottiene se l'apprendimento viene fermato prima che si verifichi un overfit dei dati di addestramento (punto P).

Figura 8.1 - Rappresentazione compatta del B.E.L..

Figura 9.1- Finestra generatore Grafici - IBM® SPSS® Statistics.

Figura 9.2 - Finestra generatore Analizza - IBM® SPSS® Statistics.

Figura 9.3 - Istogramma tipo per lo studio di frequenza - IBM® SPSS® Statistics.

Figura 9.4 - Esempio di plot di serie - GMDH Shell.

Figura 9.5 - Esempio di grafico delle frequenze - GMDH Shell.

Figura 9.6 - Esempio di Scatter - GMDH Shell.

Figura 9.7 - Esempio di matrice Scatter - GMDH Shell.

Figura 9.8 - Esempio di Scatter in 3D - GMDH Shell.

Figura 9.9 - Esempio di tabella delle statistiche - GMDH Shell.

Figura 9.10 - Esempio di istogramma descrittivo - GMDH Shell.

Figura 9.11 - Esempio di grafico delle autocorrelazioni - GMDH Shell.

Figura 9.12 - Esempio di tabella delle correlazioni - GMDH Shell.

Figura 9.13 - Esempio di diagramma 3D delle correlazioni - GMDH Shell.

Figura 9.14 - Esempio di mappa a gradienti di colore per le correlazioni - GMDH Shell.

Figura 9.15 - Esempio di mappa a gradienti di colore con curve di valori per le correlazioni - GMDH Shell

Figura 9.16 - Esempio di grafico OHLC - GMDH Shell.

Figura 9.17 - Pannello delle trasformazioni pre-elaborazione - GMDH Shell.

Figura 9.18 - Modulo risolutore - GMDH Shell.

Figura 9.19 - Finestra dei risultati - GMDH Shell.

Figura 9.20 - Finestra del modello - GMDH Shell.

Figura 10.1 – Matrice degli scatter DataSet Sicilia.

Figura 10.2 - Indici di correlazione per consumo in ktep nel residenziale.

Figura 10.3 - Indici di correlazione per consumo in ktep nei servizi non vendibili.

Figura 10.4 - Indici di correlazione per consumo in ktep in agricoltura.

Figura 10.5 - Indici di correlazione per consumo in ktep nell'industria.

Figura 10.6 - Indici di correlazione per consumo in ktep nel terziario.

Figura 10.7 - Indici di correlazione per consumo in ktep nelle attività produttive.

Figura 10.8 – Diagramma neurale - Previsione “Consumi EE regionali”.

Figura 10.9 – Plot model fit e previsione ktep residenziale Modello di previsione “Consumi EE regionali” con NN.

Figura 10.10 – Plot model fit e previsione ktep servizi non vendibili Modello di previsione “Consumi EE regionali” con NN.

Figura 10.11 – Plot model fit e previsione ktep agricoltura Modello di previsione “Consumi EE regionali” con NN.

Figura 10.12 – Plot model fit e previsione ktep industria Modello di previsione “Consumi EE regionali” con NN.

Figura 10.13 – Plot model fit e previsione ktep terziario Modello di previsione “Consumi EE regionali” con NN.

Figura 10.14 – Plot model fit e previsione ktep attività produttive Modello di previsione “Consumi EE regionali” con NN.

Figura 10.15 – Plot model fit e previsione ktep residenziale – Modello di previsione “Consumi EE regionali”.

Figura 10.16 – Plot model fit e previsione ktep servizi non vendibili – Modello di previsione “Consumi EE regionali”.

Figura 10.17 – Plot model fit e previsione ktep agricoltura – Modello di previsione “Consumi EE regionali”.

Figura 10.18 – Plot model fit e previsione ktep industria – Modello di previsione “Consumi EE regionali”.

Figura 10.19 – Plot model fit e previsione ktep terziario – Modello di previsione “Consumi EE regionali”.

Figura 10.20 – Plot model fit e previsione ktep attività produttive Modello di previsione “Consumi EE regionali”.

Figura 10.21 – Matrice degli scatter DataSet Provincie Sicilia.

Figura 10.22 - Indici di correlazione per consumo in ktep nel residenziale.

Figura 10.23 - Indici di correlazione per consumo in ktep in agricoltura.

Figura 10.24 - Indici di correlazione per consumo in ktep nell'industria.

Figura 10.25 - Indici di correlazione per consumo in ktep nel terziario.

Figura 10.26 – Diagramma neurale - Previsione “Consumi EE provinciali”.

Figura 10.27 – Plot model fit e previsione ktep residenziale Modello di previsione “Consumi EE provinciali” con NN.

Figura 10.28 – Plot model fit e previsione ktep agricoltura Modello di previsione “Consumi EE provinciali” con NN.

Figura 10.29 – Plot model fit e previsione ktep industria Modello di previsione “Consumi EE provinciali” con NN.

Figura 10.30 – Plot model fit e previsione ktep terziario Modello di previsione “Consumi EE provinciali” con NN.

Figura 10.31 – Matrice degli scatter DataSet Sicilia.

Figura 10.32 - Indici di correlazione per consumo in tep di energia elettrica.

Figura 10.33 - Indici di correlazione per consumo in tep di gas.

Figura 10.34 Correlazione tra tep energia elettrica, tep gas naturale e popolazione.

Figura 10.35 - Correlazione tra tep energia elettrica, tep gas naturale e km di rete gas.

Figura 10.36 – Diagramma neurale - Previsione “Consumi EE e gas comunali”

Figura 10.37 – Plot model fit e previsione tep energia elettrica Modello di previsione “Consumi EE e gas comunali” con NN.

Figura 10.38 – Plot model fit e previsione tep gas naturale Modello di previsione “Consumi EE e gas comunali” con NN.

Figura 10.39 – Plot residui previsione tep energia elettrica “Consumi EE e gas comunali” con NN.

Figura 10.40 – Plot residui previsione tep gas naturale “Consumi EE e gas comunali” con NN.

Figura 10.41 – Autocorrelazione previsione tep energia elettrica “Consumi EE e gas comunali” con NN.

Figura 10.42 – Autocorrelazione previsione tep gas naturale “Consumi EE e gas comunali” con NN.

Figura 10.43 – Frequenza residui previsione tep energia elettrica “Consumi EE e gas comunali” con NN.

Figura 10.44 – Frequenza residui previsione tep gas naturale “Consumi EE e gas comunali” con NN.

Figura 11.2 - Bilancio energia elettrica Sicilia – produzione e perdite.

Figura 11.3 - Produzione lorda per tipologia di impianti – Sicilia 2014 [GWh].

Figura 11.4 - Produzione lorda da termoelettrico – Sicilia 2014 [GWh].

Figura 11.5 - Produzione lorda da eolico e fotovoltaico – Sicilia 2014 [GWh].

Figura 11.6 - Perdite ed export – Sicilia 2014 [GWh].

Figura 11.7 - Consumi – Sicilia 2014 [GWh].

Figura 11.8 - Produzione – Sicilia 2014 [GWh].

Figura 11.9 - Numero impianti FER – Italia 2014.

Figura 11.10 - Potenza impianti FER – Italia 2014 [MW].

Figura 11.11 - Produzione impianti FER – Italia 2014 [GWh].

Figura 11.12 - Produzione impianti FER per superficie territoriale – Italia 2014 [MWh/km<sup>2</sup>].

Figura 11.13 - Numero ed incremento impianti FER – Sicilia.

Figura 11.14 - Potenza annuale ed incremento impianti FER – Sicilia [MWh].

Figura 11.15 - Produzione annuale ed incremento impianti FER – Sicilia.

Figura 11.16 - Potenza impianti FER – Sicilia [MW].

Figura 11.17 - Produzione impianti FER – Sicilia [GWh].

Figura 11.18 - Matrice degli scatter del DataSet.

Figura 11.19 - Matrice degli scatter del DataSet.

Figura 11.20 - Indici di correlazione per PV tep.

Figura 11.21 - Indici di correlazione per Eolico tep.

Figura 11.22 - Indici di correlazione per Idro fl tep.

Figura 11.23 - Correlazione tra PV tep ed indici climatici medi.

Figura 11.24 - Correlazione tra Eolico tep ed indici climatici medi.

Figura 11.25 - Correlazione tra Idro fl tep ed indici climatici medi.

Figura 11.26 – Diagramma neurale - Previsione “Produzione EE da FER”.

Figura 11.27 – Plot model fit e previsione PV tep– Modello di previsione “Produzione EE da FER” con NN.

Figura 11.28 – Plot model fit e previsione Eolico tep– Modello di previsione “Produzione EE da FER” con NN.

Figura 11.29 – Plot model fit e previsione Idro fl tep– Modello di previsione “Produzione EE da FER” con NN.

Figura 11.30 – Plot residui previsione PV tep– Modello di previsione “Produzione EE da FER” con NN.

Figura 11.31 – Plot residui previsione Eolico tep– Modello di previsione “Produzione EE da FER” con NN.

Figura 11.32 – Plot residui previsione Idro fl tep– Modello di previsione “Produzione EE da FER” con NN.

Figura 11.33 – Autocorrelazione previsione PV tep– Modello di previsione “Produzione EE da FER” con NN.

Figura 11.34 – Autocorrelazione previsione Eolico tep– Modello di previsione “Produzione EE da FER” con NN.

Figura 11.35 – Autocorrelazione previsione Idro fl tep– Modello di previsione “Produzione EE da FER” con NN.

Figura 11.36 – Frequenza residui previsione PV tep– Modello di previsione “Produzione EE da FER” con NN.

Figura 11.37 – Frequenza residui previsione Eolico tep– Modello di previsione “Produzione EE da FER” con NN.

Figura 11.38 – Frequenza residui previsione Idro fl tep– Modello di previsione “Produzione EE da FER” con NN.

Figura 11.39 – Plot model fit e previsione PV tep “Produzione EE da FER”.

Figura 11.40 – Plot model fit e previsione eolico tep “Produzione EE da FER”.

Figura 11.41 – Plot model fit e previsione Idro fl tep “Produzione EE da FER”.

Figura 12.1 - Frequenza valori in tep degli Acquisti dal 2011 al 2014.

Figura 12.2 - Frequenza valori in tep delle Vendite dal 2011 al 2014.

Figura 12.3 - Frequenza valori in €/tep dei Prezzi dal 2011 al 2014.

Figura 12.4 - Frequenza valori in tep dei delta dal 2011 al 2014.

Figura 12.5 - Frequenza valori in tep della produzione oraria da idroelettrico dal 2011 al 2014.

Figura 12.6 - Frequenza valori in tep della produzione oraria da fotovoltaico dal 2011 al 2014.

Figura 12.7 - Frequenza valori in tep del consumo orario dei pompaggi dal 2011 al 2014.

Figura 12.8 - Frequenza valori in tep della produzione oraria da termoelettrico dal 2011 al 2014.

Figura 12.9 - Frequenza valori in tep della produzione oraria da eolico dal 2011 al 2014.

Figura 12.10 - Potenza installata in MW per singola fonte dal 2011 al 2014.

Figura 12.11 - Frequenza dei ktep di acquisto nel MGP per gli anni 2011,2012, 2013 e 2014.

Figura 12.12 - Frequenza dei ktep di vendite nel MGP per gli anni 2011,2012, 2013 e 2014

Figura 12.13 - Frequenza dei prezzi in €/tep d nel MGP per gli anni 2011,2012, 2013 e 2014.

Figura 12.14 - Frequenza del parametro delta in tep nel MGP per gli anni 2011,2012, 2013 e 2014.

Figura 12.15 - Plot produzione idroelettrica dal 2011 al 2014.



- Figura 12.16 - Plot produzione fotovoltaica dal 2011 al 2014.
- Figura 12.17 - Plot energia richiesta dagli impianti a pompaggio dal 2011 al 2014.
- Figura 12.18 - Plot produzione termoelettrica dal 2011 al 2014.
- Figura 12.19 - Plot produzione eolica dal 2011 al 2014
- Figura 12.20 - Frequenza Delta-Prezzi nulli
- Figura 12.21 - Frequenza ore a prezzo nullo.
- Figura 12.22 – Diagramma neurale - Previsione “Prezzo Zonale”.
- Figura 12.23 – Plot model fit e previsione “Prezzo Zonale” con Neural Network.
- Figura 12.24 – Plot residui previsione “Prezzo Zonale” con Neural Network.
- Figura 12.25 – Autocorrelazione previsione “Prezzo Zonale” con Neural Network.
- Figura 12.24 – Frequenza residui previsione “Prezzo Zonale” con Neural Network.
- Figura 12.25 – Plot model fit e previsione “Prezzo Zonale”
- Figura 13.1 - Classificazione dei climi europei.
- Figura 13.2 - HDD, CDD e radiazione solare degli Stati europei.
- Figura 13.3 - Paesi oggetto dello studio e relativi valori di HDD e CDD.
- Figura 13.3 - HDD massimo e minimo degli stati oggetto dello studio.
- Figura 13.4 - Planimetria tipica di edificio tipo ufficio.
- Figura 13.5 - Rendering dei prospetti dell'edificio tipo ufficio.
- Figura 13.6 - Rappresentazione delle zone termiche dell'edifici tipo ufficio.
- Figura 13.7 - Modello costruito su TRNSYS.
- Figura 13.8 - Correlazione tra HDD e domanda termica.
- Figura 13.9 - Correlazione tra HDD e domanda termica per Belgio e Germania.
- Figura 13.10 - Correlazione tra HDD e domanda termica per Spagna e Francia.
- Figura 13.11 - Correlazione tra HDD e domanda termica per Italia e Svezia.
- Figura 13.12 - Correlazione tra HDD e domanda termica per Regno Unito
- Figura 13.13 - Correlazione tra HDD e domanda termica per il nord Europa.
- Figura 13.14 - Correlazione tra HDD e domanda termica per il centro Europa.
- Figura 13.15 - Correlazione tra HDD e domanda termica per il sud Europa.



## CAPITOLO 1 *Premessa*

Le azioni internazionali per la riduzione delle emissioni, l'indipendenza energetica, l'efficienza degli impianti e del sistema edificio, costituiscono le basi della politica energetica di ogni Stato industrializzato. Tali temi risultano strategici anche perché trasversali ad altri, quali: la libera concorrenza e competitività delle imprese, la crescita economica, il benessere dei cittadini, etc..

Il sistema energetico europeo è stato al centro di diverse delibere e decisioni da parte degli organismi preposti dell'Unione Europea (U.E.) che lo hanno portato, di fatto, all'odierna configurazione. Sarà tuttavia nuovamente soggetto a mutare, al fine di completare il percorso che ebbe inizio con la nascita della C.E.C.A e dell'ERATOM e che è proseguito con la sottoscrizione del protocollo di Kyoto e l'emanazione del pacchetto Clima ed Energia 20-20-20.

Nonostante i numerosi sforzi intrapresi, ancora oggi, l'U.E. rappresenta il primo importatore di energia al mondo (53% del proprio fabbisogno con un costo di circa 400 miliardi di euro all'anno) e ben 12 Stati membri non soddisfano l'obiettivo minimo di interconnessione elettrica, mentre 6 Stati dipendono da un unico fornitore esterno per tutte le loro importazioni di gas. Il tutto causa, di fatto, prezzi all'ingrosso dell'elettricità e del gas più elevati, rispettivamente del 30% e del 100% rispetto a quelli praticati negli U.S.A. [1]. Nel contesto degli edifici, il 75% del parco immobiliare U.E. è a bassa efficienza energetica; mentre nei trasporti il 94% dipende dai prodotti petroliferi, di cui il 90% importati [2].

Partendo da questi dati si stima che solo nel periodo fino al 2020 occorrerà investire oltre 1.000 miliardi di euro nel settore dell'energia. Tuttavia, le decisioni prese nel recente passato hanno generato una diminuzione delle emissioni di gas a effetto serra nell'U.E. pari al 18% nel periodo 1990-2011; mentre nel comparto delle energie rinnovabili si è ottenuto un fatturato annuo di 129 miliardi di euro per le imprese dell'U.E. dando lavoro a più di un milione di addetti [3].

Le sfide che l'U.E., ed i suoi Stati membri, si sono impegnati ad intraprendere nei prossimi anni consistono nel conservare il ruolo guida dell'Europa negli investimenti globali per le energie rinnovabili e realizzare una rete europea dell'energia adeguatamente interconnessa che, secondo stime, potrebbe generare risparmi fino a 40 miliardi di euro l'anno per i consumatori. Per far ciò l'U.E. ha preso la decisione di ridurre entro il 2030 le emissioni di gas a effetto serra almeno del 40%, incrementare l'energia da fonti rinnovabili almeno del 27% e migliorare l'efficienza energetica almeno del 27% [4].

Tale decisione avrà senza alcun dubbio una ricaduta locale sui singoli Stati, ed a secondo della propria *governance* del territorio, sui livelli di governo sub-nazionali (per lo stato italiano: regioni, città metropolitane, liberi consorzi di comuni).

Da qui l'importanza che le principali strutture di *governance* del territorio si dotino di strumenti programmatori e pianificatori che aiutino il decisore nel perseguire le migliori scelte legislative e regolatorie per l'intero sistema energetico; soprattutto alla luce dell'inerzia del sistema stesso, a livello locale e nazionale, a mutare la propria configurazione su vasta scala, come è ampiamente documentato dai risultati ottenuti post attività legislative e regolatorie del passato ventennio. Infatti, odierni studi hanno evidenziato come in Italia gli interventi legislativi e regolatori per le fonti energetiche rinnovabili e l'efficienza energetica, trovano fattivo riscontro a due anni dalla loro emanazione, completando il loro scopo nei successivi cinque anni [5].

Non di meno importante è il tema della dinamicità per gli strumenti programmatori. Infatti, la costante ricerca tecnologica ed i mutamenti del quadro politico ed economico internazionale possono indurre alla necessità di variare le scelte di politica energetica al fine di salvaguardare l'economia e la competitività del singolo Stato. Alla base di tutti gli strumenti c'è naturalmente il dato energetico, le serie storiche e la loro iterazione con indicatori afferenti ad altre sfere, sempre trasversali all'energia.

Il lavoro di ricerca nell'ambito del dottorato, svolto da gennaio 2013 a dicembre 2015, ha quindi riguardato l'implementazione ed analisi di modelli numerici, opportunamente costruiti attraverso l'individuazione di specifici indicatori energetici e di opportuni fattori correttivi, che determinano in maniera predittiva e/o statistica i quantitativi di energia producibile da fonti non programmabili e di energia consumata per vettore e settore produttivo al sussistere di determinate condizioni al contorno. Nel corso delle diverse fasi della ricerca, valorizzando il periodo trascorso all'estero, sono stati implementati ulteriori casi studio che riguardano il mercato elettrico e i consumi in edifici con caratteristiche di ufficio siti in diversi stati europei.

Utilizzando come caso studio la Regione Sicilia, grazie alla collaborazione con diversi enti istituzionali si è recuperato un DataBase energetico vasto, sia per tipologia di dato che per scala temporale. Sono stati quindi sviluppati per i differenti casi studio i modelli predittivi basati su *Neural Network* ed, ove possibile, anche quelli basati su *Serie Storiche*. Nel complesso i modelli implementati sia con *Neural Network* che attraverso *Serie Storiche* hanno mostrato valori contenuti di errori sulle previsioni (non superiori al 20%). Per DataSet costituiti da un numero basso di dati, i modelli basati su *Serie Storiche* sono risultati superiori per affidabilità rispetto ai modelli *Neural Network* anche se, in questo caso, la predizione assume più una veste di predizione del trend che del dato effettivo. Con DataSet ben organizzati i risultati per modelli regressivi *Neural Network* risultano essere più affidabili e precisi, fornendo indici di bontà quanto più elevati al crescere del numero di campioni del DataSet.

## **CAPITOLO 2** *Le fasi della ricerca*

L'attività di ricerca ha affrontato differenti ambiti tipici della pianificazione energetica, dell'analisi di dati e dei modelli predittivi. Tutti i predetti temi sono coniugati al fine di studiare le caratteristiche dei sistemi energetici, i bilanci energetici a scala regionale e locale, di valutare gli indicatori di efficienza dei sistemi su macro area utilizzando metodi statistici e previsionali. A tal fine sono stati applicati modelli e software predittivi basati su metodi di regressione e/o Neural Network per la predisposizione di scenari energetici a breve, medio e lungo termine.

Nel complesso, il progetto di ricerca ha riguardato l'implementazione ed analisi di modelli numerici, in grado di determinare in maniera predittiva e/o statistica i quantitativi di energia producibile da fonti non programmabili e di energia consumata per vettore e settore produttivo. Nel corso del periodo trascorso all'estero, sono stati altresì implementati ulteriori casi studio che riguardano il mercato elettrico e i consumi in edifici con caratteristiche di ufficio siti in diversi stati europei. Pertanto, la ricerca condotta ha previsto quattro fasi, di seguito esposte.

### **2.1 Step 1 – Ricerca bibliografica e normativa di settore**

In primo luogo sono state svolte delle accurate ricerche bibliografiche per avere contezza dello stato della ricerca scientifica, anche internazionale, sul tema dell'analisi dei dati energetici e dello sviluppo di scenari.

Una attenta ricognizione dei risultati delle ricerche scientifiche condotte nei dieci anni precedenti ha quindi permesso di accertare lo stato dell'arte sul tema dell'analisi dei dati energetici e della proiezioni di scenari. Vi è un ampio numero di pubblicazioni scientifiche aventi ad oggetto la previsione della domanda di energia su macro area, ed in particolare di elettricità, sia a breve che a lungo termine. In generale, i metodi applicati agli studi esistenti possono essere suddivisi in tre categorie:

1. Metodi di previsione semplici e tradizionali;
2. Modelli di previsione econometrici;
3. Metodi di previsione evolutivi basati sulle tecniche di intelligenza artificiali.

Sono stati quindi svolti approfondimenti relativi alla legislazione e alle regole nazionali ed internazionali in materia di monitoraggio dei dati energetici, al fine di acquisire informazioni corrette e standardizzate circa la natura dei dati, i metodi di analisi ed elaborazione e gli errori di stima presenti

nelle serie storiche. Approfondimenti a livello di regolazione mondiale hanno portato lo studio all'identificazione delle principali fonti normative concernenti le metodologie di analisi statistica di dati energetici, che sono:

1. Concepts and methods in energy statistics, with special reference to energy accounts and balances: a technical report, Studies in Methods, Series F, No. 29, United Nations, New York, 1982;
2. Energy statistics: definitions, units of measure and conversion factors, Studies in Methods, Series F, No. 44, United Nations, New York, 1987;
3. Energy statistics: a manual for developing countries, Studies in Methods, Series F, No. 56, United Nations, New York, 1991;
4. Energy Statistics Manual, OECD/IEA/EUROSTAT, Paris, 2004.

A livello europeo, al fine di monitorare l'impatto e le conseguenze della politica dell'U.E. in campo energetico, è stata evidenziata con più provvedimenti la necessità di dati precisi e tempestivi sulle quantità di energia, sulle forme, sulle fonti, sulla generazione, sull'approvvigionamento, sulla trasformazione e sui consumi. Pertanto il regolamento n. 1099/2008 relativo alle statistiche dell'energia, adottato dall'U.E. e recepito da tutti gli stati membri, ha istituito un quadro comune per la produzione, la trasmissione, la valutazione e la diffusione di statistiche dell'energia comparabili nella Comunità.

A livello nazionale i provvedimenti più significativi sono stati:

1. il d.M. del 14 gennaio 2012 del Ministero dello Sviluppo economico, che ha approvato la metodologia che, nell'ambito del sistema statistico nazionale in materia di energia, è applicata per rilevare i dati necessari a misurare il grado di raggiungimento degli obiettivi nazionali in materia di quote dei consumi finali lordi di elettricità, energia per il riscaldamento e il raffreddamento, e per i trasporti coperti da fonti energetiche rinnovabili.
2. il d.M. del 11 maggio 2015 del Ministero dello Sviluppo economico, di approvazione della metodologia che, nell'ambito del sistema statistico nazionale, è applicata per rilevare i dati necessari a misurare il grado di raggiungimento degli obiettivi regionali, in attuazione dell'articolo 40, comma 5, del decreto legislativo 3 marzo 2011, n. 28.

## **2.2 Step 2 – Metodi di analisi e predizione dei dati storici**

Ai fini di sviluppare i temi della ricerca attraverso un approccio scientifico, è stata svolta una attenta analisi dei metodi statistici per l'analisi dati e dei metodi predittivi per la creazione di scenari, con particolare interesse a quelli più idonei a trattare serie storiche.

Sono quindi stati approfonditi i seguenti argomenti e metodi:

1. Definizione di dati storici;
2. I tipi di analisi grafiche e preliminari;
3. I tipi di analisi descrittive dei dati;
4. Le fasi delle analisi di predizione;
5. I metodi di trasformazione ed aggiustamento del dato;
6. I metodi di predizione qualitativi e quantitativi e gli approcci classici e moderni;
7. I metodi informali;
8. I metodi di scomposizione;
9. I metodi di lisciamento esponenziale;
10. I metodi di regressione.

Successivamente, sono stati approfonditi i modelli più comuni per l'analisi dati e realizzazione di forecasting, quali: i modelli ARIMA (AutoRegressive Integrated Moving Average) e GMDH (Group Method of Data Handling). È stato inoltre svolto un approfondimento sui metodi di previsione intelligenti ed artificiali, quali le Reti-Neurali (Neural Network). Sulla base degli studi svolti, degli approfondimenti condotti e dei dati raccolti, sono stati identificati ed approfonditi i software che meglio si sono prestati agli scopi del lavoro. La ricerca ha quindi condotto ad utilizzare:

- il pacchetto IBM SPSS, che è un software di analisi, classificazione e clustering di dati sperimentali;
- il pacchetto GMDH Shell (Forecasting Software for Business), che è un software di analisi, classificazione e Clustering di dati sperimentali che realizza attraverso algoritmi complessi o reti-neurali delle analisi previsionali.

Gli studi condotti hanno infine determinato la realizzazione di numerosi modelli all'interno dei predetti software.

## **2.3 Step 3 – Reperimento ed organizzazione del dato energetico**

Una volta acquisita contezza del panorama di ricerca internazionale e nazionale in materia di dati energetici, delle tecniche statistiche e di forecasting più comuni, della tipologia di dati trattati e delle regole e metodologie che stanno alla base delle banche dati internazionali e nazionali, è stata avviata la fase di ricognizione dei dati e di implementazione di un DataSet funzionale alle analisi dei casi studio da implementare nel lavoro di ricerca.

In primo luogo è stato svolto un approfondimento sulle tipologie di dati energetici disponibili, sulle unità di misura utilizzate nelle statistiche, sui dati socio-economici ed ambientali che pesano sulle dinamiche energetiche locali, sugli indici energetici ed ambientali maggiormente utilizzati per descrivere i sistemi energetici locali ed i bilanci energetici. I dati sono stati quindi recuperati dagli attori principali del sistema energetico italiano e da chi ha ruoli di Governance dello stesso. In particolare, in Italia, il sistema energetico è organizzato in tre differenti step: produzione, trasporto/dispacciamento, vendita/consumo. L'implementazione del DataSet è frutto di un intenso lavoro di ricognizione, disaggregazione ed organizzazione dei dati forniti a livello nazionale e regionale da: Ministero dello Sviluppo Economico, Gestore Sistemi Energetici, E.N.E.A, TERNA S.p.a., ENEL, Assopetroli, Assocostieri, Agenzia delle Dogane, Assessorato regionale dell'Energia e dei Servizi di Pubblica Utilità della Regione Siciliana.

Oltre ai dati strettamente energetici sono stati raccolti, elaborati ed organizzati dati socio-economici provenienti da Eurostat, Istat ed Istituto Guglielmo Tagliacarne.

La ricognizione dei dati meteorologici è avvenuta con l'ausilio di: report e documenti provenienti da enti specifici; di DataBase relativi a stazioni meteo distribuiti sul territorio della regione oggetto del caso studio; di indici espressi in metodologie nazionali approvate dai Ministeri competenti e di dati pubblicati da Eurostat. Tutti i dati sono stati quindi organizzati, disaggregati ove necessario, ed allineati l'un l'altro in termini temporali, al fine di poterli confrontare e analizzare la sussistenza di eventuali correlazioni tra essi.

## **2.4 Step 4 – Implementazione dei casi studio**

Acquisita la conoscenza dei software scelti per le analisi ed organizzato il DataSet, sono stati implementati i casi studio che possono essere distinti in quattro macro aree:

1. Consumi energetici a scala regionale e locale;
2. Produzione di energia elettrica da FER;
3. Prezzo dell'energia elettrica;
4. Consumi energetici in edifici uso ufficio.



I casi studio delle prime tre aree riguardano la Regione Sicilia. La prima area è stata implementata attraverso l'analisi e le proiezioni dei consumi di energia elettrica e combustibili gassosi nei settori: residenziale, terziario, industria ed agricoltura.

Per quanto riguarda l'area della produzione di energia elettrica da FER, attraverso i dati di produzione oraria per singola fonte e del livello di domanda richiesta, sono state verificate le interazioni dei valori registrati con variabili ambientali e sono stati implementati modelli di predizione dei valori di produzione.

La terza area ha visto l'analisi del comportamento del mercato elettrico al variare della potenza ed energia prodotta dalle FER-E. È stato altresì sviluppato un modello predittivo che correla il prezzo dell'energia al valore di produzione e potenza degli impianti rinnovabili presenti in una data zona.

Il quarto caso studio si basa sulla ricerca condotta presso lo Zentrum für nachhaltige Energietechnik – Institut für Angewandte Forschung (IAF) der Hochschule für Technik Stuttgart, che ha visto la raccolta, ed organizzazione di dati relativi al consumo di edifici uso ufficio nel contesto europeo. Sono state quindi elaborate una serie di correlazioni numeriche che legano gli indicatori climatici che caratterizzano i differenti climi europei al consumo termico di edifici ad uso ufficio. I risultati sono stati verificati con dati rilevati in campo provenienti da differenti pubblicazioni scientifiche internazionali.

## CAPITOLO 3 *Ricerche bibliografiche*

Una attenta ricognizione dei risultati delle ricerche scientifica ha permesso di apprezzare lo stato dell'arte sul tema dell'analisi dei dati energetici e della proiezioni di scenari. Vi è un ampio numero di pubblicazioni scientifiche ad oggetto la previsione della domanda di energia, ed in particolare di elettricità, sia a breve che a lungo termine. In generale, gli studi esistenti possono essere suddivisi in tre categorie.

### *1. Metodi di previsione semplici e tradizionali.*

Le prime pubblicazioni in materia hanno visto l'elaborazione di alcune espressioni matematiche per prevedere la domanda di energia elettrica attraverso l'analisi di alcuni indicatori economici. Le metodologie predittive più comuni sono basate su il *coefficiente di elasticità elettrica* ed il *fattore di consumo per abitante*. La facilità di implementazione e la semplicità applicativa rendono questi metodi molto utilizzati. Tuttavia, essi si concentrano principalmente sulla previsione di carico. Inoltre, la loro applicabilità è limitata a ben definite condizioni, che se non soddisfatte forniscono un grado di precisione della previsione basso.

### *2. Modelli di previsione econometrici.*

Un approccio di questo tipo include metodi basati sulla regressione, modelli di serie temporali, metodo *State-Space* e filtro *Kalman*. Tra questi, i modelli più utilizzati per le previsioni dei consumi energetici sono quelli di regressione; questi modelli stabiliscono il rapporto tra domanda e fattori esterni quali il PIL o al PIL pro capite, la popolazione, e il tempo. A tal proposito metodi basati sulla regressione lineare multipla sono alla base di molti studi di previsione della domanda di energia. Un inconveniente di questi modelli è dato dall'alto condizionamento dei dati di output da parte dei fattori di input, ed in particolare dalla loro completezza e affidabilità. A tal fine deve essere posta maggiore attenzione nella raccolta dei dati energetici e nella realizzazione del DataSet.

Al contrario, i modelli basati su serie storiche richiedono solo i dati storici dei dati di output, e proprio per questo sono anch'essi molto utilizzati per le previsioni della domanda di energia. Tra i differenti modelli di serie storiche, l'auto-regressivo integrata a media mobile (ARIMA) è il modello più comunemente usato ed è ampiamente applicato. Tuttavia, questo tipo di modello non può affrontare problemi multi-variabili.

In ultimo, il modello *State-Space* e filtro *Kalman*, in cui l'output è modellato come una variabile di stato utilizzando formulazione *State-Space*, è comune, ma meno utilizzato dei precedenti, per le previsioni dei consumi energetici.

### 3. *Metodi di previsione evolutivi basati sulle tecniche di intelligenza artificiali.*

Negli ultimi due decenni, metodi basati su Artificial Intelligence (AI) hanno dimostrato la capacità di fornire migliori risultati nell'elaborazione di modelli non lineari ed ad alta complessità. Questi metodi hanno preso sempre più piede per le previsioni della domanda energetica in quanto non richiedono alcuna formulazione matematica complessa o correlazione quantitative tra input ed output.

Partendo da quanto prima esposto, si è svolta un'attenta ricognizione su un arco temporale di 15 anni di prodotti scientifici, non andando oltre l'anno 2000 poiché si è ritenuto che questo arco temporale possa essere più che esaustivo per caratterizzare le tecniche e le analisi che contraddistinguono l'odierno panorama scientifico, rappresentando al meglio le tecniche ormai oggi più consolidate per le previsioni dei consumi energetici.

Si riportano di seguito un breve sunto degli aspetti più importanti delle pubblicazioni ritenute utili ai fini della ricerca in oggetto e che tracciano sia la direzione che ha assunto la ricerca ma anche il punto di arrivo che ne consegue sul tema dell'analisi dei dati energetici e dell'elaborazione degli scenari.

Dalla pubblicazione *The trend in current and near future energy consumption from a statistical perspective* [6], si evince, come affermato dagli stessi autori, che i principali fattori da cui dipende il consumo energetico sono quattro:

1. aumento della popolazione;
2. miglioramento del tenore di vita;
3. sviluppo di tecnologie scientifiche;
4. condizioni di ogni singolo paese (risparmio energetico, struttura industriale, stile di vita, etc.).

La definizione di questi fattori è complessa, anche per il fatto che sono interdipendenti tra loro.

Nel dettaglio, l'influenza del miglioramento del tenore di vita sarebbe dimostrata dal PIL pro-capite (Prodotto Interno Lordo) (PIL / popolazione). Mentre l'influenza dello sviluppo tecnologico e delle condizioni di ogni singolo paese hanno come indicatore: il consumo di energia sul PIL (consumo di energia / PIL), in quanto fornisce la quantità di energia consumata per ogni prodotto di valore aggiunto.

Pertanto gli autori affermano che, se si parte dal presupposto che la correlazione tra consumo energetico e i fattori sopra riportati continuerà nel prossimo futuro, la domanda di energia (consumo di energia) può essere stimata dalle previsioni demografiche, dalle previsioni dei tassi di crescita del PIL, e dalle rispettive curve di correlazione ricavate

È altresì affermato dagli stessi autori, che è scientificamente accettato che, pro-capite, il consumo di energia aumenta proporzionalmente all'aumento del PIL pro-capite. Inoltre, è dimostrato che il rapporto di proporzione è più grande, quindi ha un maggiore correlazione, per quegli stati in cui il PIL pro-capite è piccolo (stati in via di sviluppo), mentre diventa più piccolo nei paesi avanzati in cui il PIL pro-capite è di grandi dimensioni.

Dalle elaborazioni svolte, è quindi comprovato come nei paesi più avanzati l'influenza della popolazione sul consumo energetico peserà sempre meno, mentre maggiore importanza avrà la struttura del consumo energetico (miglioramento del tenore di vita, sviluppo di tecnologie scientifiche, risparmio energetico, struttura industriale, stile di vita, etc.).

Dall'approfondimento della pubblicazione *Electricity estimation using genetic algorithm approach: a case study of Turkey* [7], si evince come la stima della domanda di energia elettrica può avvenire: con varie forme di espressioni matematiche che vedono come variabili alcuni indicatori socio-economico; con *Fuzzy Logic*; con Neural Networks o altri metodi. In particolare, le espressioni matematiche possono essere lineari o non lineari. A causa della fluttuazione degli indicatori socio economici, un'equazione non lineari tende a meglio stimare il futuro fabbisogno di energia elettrica. A tal fine risulta interessante stimare se un approccio che vede l'utilizzazione di algoritmi genetici possa ottimizzare le tecniche di previsione dei consumi energetici, come applicato con successo su una vasta gamma di problemi di ottimizzazione ingegneristica.

Nella pubblicazione *Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks* [8], gli autori presentano tre tecniche di modellazione per la previsione dei consumi energetici di energia elettrica: analisi di regressione tradizionale, albero di decisione e Neural Networks.

Tradizionalmente le tecniche di modellazione basate su analisi di regressione sono state le più popolari per predire i consumo energetico. Le Neural Networks negli ultimi anni si stanno affermando come strumento affidabile per la stima previsionale del consumo di energia. La metodologia basata sugli alberi di decisione è invece risultata uno strumento efficiente per la modellizzazione dei sistemi di produzione.

Mentre il metodo di analisi di regressione è supportata da teorie statistiche e fornisce buone stime, ciò ancora non è per le tecniche basate su albero di decisione e Neural Networks. Negli ultimi dieci anni, progressi nella gestione di Database e miglioramenti in termini di velocità di calcolo hanno portato a

nuovi modi di condurre l'analisi dei dati. Lo studio condotto dagli autori illustra che il modello albero decisionale e il modello Neural Networks forniscono risultati leggermente migliori rispetto al modello tradizionale di analisi di regressione, rispettivamente: il primo nel periodo estivo ed il secondo nel periodo invernale. In generale, le differenze fra i risultati forniti dai tre tipi di modelli sono abbastanza piccole, pertanto le tre tecniche di modellazione sono generalmente paragonabili nel predire il consumo di energia.

La ricerca esposta in *Short-term prediction of household electricity consumption: Assessing weather sensitivity in a Mediterranean area*. [9], rileva che nel settore residenziale, il consumo di energia dipende dalle caratteristiche dell'involucro edilizio e sul comportamento degli occupanti. Quest'ultimo è soggetto a molti fattori tra cui le preferenze di comfort soggettive e le loro caratteristiche socio-demografiche e socio-economiche. D'altra parte, è noto che consapevolezza e atteggiamenti verso il consumo di energia sono più evidenti nel consumo domestico che in situazioni in cui molte persone possono contemporaneamente avere un'influenza sul consumo di energia, come ad esempio in uffici. Oltre alla molteplicità di fattori umani che influenzano il fabbisogno energetico, numerosi studi hanno dimostrato l'influenza delle variabili climatiche sul consumo di energia e soprattutto sulla domanda di energia elettrica. Gli approcci finora utilizzati per la previsione del carico includono modelli statistici/regressione o modelli econometrici, modelli di ingegneria, modelli strutturali, logica fuzzy e sistemi con intelligenza artificiale.

Gli autori hanno quindi elaborato un modello previsionale basato sulle Reti Neurali (Neural Networks) che evidenzia l'importanza di un indice di comfort per la valutazione del consumo di elettricità delle famiglie, sottolineando come la domanda di elettricità interna complessiva è influenzata dall'uso di apparecchiature *Heating, Ventilating and Air Conditioning* (HVAC). Il lavoro ha quindi comprovato la robustezza di un modello predittivo basato sulle Reti Neurali per le stime di consumi energetici, questo perché il modello ha una buona capacità di mantenere prestazioni soddisfacenti anche con dati rumorosi e incompleti come quelli utilizzati nel caso studio.

Studi che integrano l'Algoritmo genetici (GA) con algoritmi integrati o le Neural Networks (NN) per stimare e prevedere la domanda di energia elettrica sono stati esposti negli articoli *Integration of genetic algorithm, computer simulation and design of experiments for forecasting electrical energy consumption* [10] e *Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption* [11]. Le elaborazioni svolte forniscono ottimi risultati per la previsione dei consumi energetici e soprattutto con applicazioni a serie storiche ai modelli integrati GA ed NN.

Elementi fondamentali per le elaborazioni sui consumi energetici sono dati dal tipo di indici utilizzati per le indagini, gli indici sono infatti necessari per caratterizzare il dato rispetto ad input esterni al sistema energetico.

Proprio sul tema degli indici, nell'articolo *Strengths and weaknesses of common sustainability indices for multidimensional systems* [12], gli autori affermano che la sostenibilità sta rapidamente passando da un concetto astratto ad uno stato misurabile di sistemi umano-ecologici dinamici. Il gran numero di indicatori economici, sociali, ed ambientali attualmente disponibili offrono una visione alle volte confusa della sostenibilità del sistema. Per facilitare le decisioni politiche, gli indicatori devono essere ben inseriti in un quadro concettuale e aggregati in indici. Tuttavia, le decisioni politiche possono essere inefficaci o addirittura controproducenti se non considerano i fattori che influenzano il comportamento degli indici: la scala dei dati disponibili e la scelta dei limiti del sistema; l'inclusione, la trasformazione e ponderazione dei dati dell'indicatore; e il metodo di aggregazione utilizzato. Quindi può succedere, come è stato già dimostrato in più pubblicazioni, che gli indici di sostenibilità non riescono a classificare in modo coerente i paesi oggetto di studi, anche quando si utilizza gli stessi indicatori. Pertanto, chi sviluppa indici di sostenibilità dovrebbe indicare in maniera chiara i limiti dello stesso indice.

Per la complessità del sistema industriale, modelli previsionali sul consumo industriale possono richiedere un approccio basato su sistemi all'avanguardia ed intelligenti come le Neural Networks, come affermato in *Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors* [13]. La necessità di modelli altamente sviluppati ed intelligenti è dovuta dal fatto che un sistema industriale è caratterizzato da diverse tipologie di industrie con attività altamente energivore ma anche con settori a bassa intensità. Studi precedenti hanno dimostrato l'affidabilità dell'approccio con Neural Networks per prevedere il consumo a breve termine. Gli autori hanno quindi dimostrato che le Neural Networks possono essere utilizzate per previsioni a lungo termine, come la previsione annuale, nonostante l'eterogeneità e forti variazioni della richiesta di energia delle industrie ad alta intensità energetica. Lo studio ha in fine dimostrato il vantaggio di un approccio con Neural Networks attraverso il confronto con i dati reali e un modello di regressione convenzionale.

Nell'articolo *A hybrid self-adaptive Particle Swarm Optimization–Genetic Algorithm–Radial Basis Function model for annual electricity demand prediction* [14], è proposto un modello ibrido di Neural Networks per la previsione della domanda annuale di elettricità. I risultati dello studio hanno evidenziato la bontà del modello ibrido, caratterizzato altresì da una struttura più semplice, superando di fatto i problemi relativi all'apprendimento fino ad oggi presenti nei modelli tradizionali e fornendo risultati con una maggiore precisione di stima rispetto ad altri modelli NN tradizionali.

Ulteriori approfondimenti su differenti metodi sono stati svolti da *An application of multivariate statistical techniques to partial equilibrium models outputs: The analysis of the NEEDS-TIMES Pan European model results* [15]. Nel documento gli autori affermano che lo sviluppo sostenibile richiede strumenti di analisi per valutare, attraverso un approccio globale, l'efficacia delle politiche energetico-

ambientali a medio e lungo termine e il loro impatto sui diversi settori macroeconomici. Tra gli strumenti a disposizione, modelli di equilibrio parziali sono particolarmente adatti a rappresentare e analizzare sistemi energetici complessi con un elevato dettaglio tecnologico, nonché di individuare le ottimali roadmap di implementazione dei sistemi energetici che permette di soddisfare molteplici obiettivi (ad esempio, la sicurezza dell'approvvigionamento energetico, mitigazione dei cambiamenti climatici e miglioramento della qualità dell'aria). In realtà questi modelli consentono agli utenti di eseguire analisi di scenario, al fine di esplorare il comportamento del sistema energetico al variare delle condizioni al contorno. Essi sono in genere caratterizzati da una enorme quantità di dati il cui contenuto informativo è spesso non pienamente sfruttato. In questo contesto, le tecniche statistiche multivariate possono rappresentare uno strumento fondamentale per caratterizzare struttura di correlazione dei dati e per caratterizzare le variabili con il più alto contenuto di informazioni. La metodica utilizzata dagli autori ha quindi permesso di ridurre la dimensione dei problemi nella caratterizzazione della struttura di correlazione dei dati, individuando le variabili più significative da un punto di vista statistico.

A conclusione del lavoro di ricerca bibliografico è possibile quindi affermare l'ormai prevalenza di studi e ricerche incentrate sulle Neural Networks che prefigurano la possibilità di superamento di tutte quelle difficoltà riscontrate nei modelli di regressione tradizionali e serie storiche, che negli anni passati hanno costituito i modelli più utilizzati nella costruzione di scenari energetici per l'elaborazione delle azioni e decisioni che hanno portato il sistema energetico europeo e nazionale all'odierna configurazione.

## **CAPITOLO 4** ***Le statistiche energetiche*** ***evoluzione della normativa di riferimento***

### **4.1 Contesto mondiale**

Il ruolo che l'energia gioca nello scenario mondiale, ed in particolare nello sviluppo socio-economico dei singoli paesi, ha fin da subito posto la necessità da parte degli organi di Governo di possedere statistiche dell'energia complete e di alta qualità.

Fin dalla sua nascita la commissione statistica delle Nazioni Unite ha discusso e deliberato su questioni relative alle statistiche dell'energia. Ed in seguito alla crisi energetica dei primi anni '70, le statistiche energetiche assumono maggiore importanza, tanto da costituire, da lì in avanti, tema speciale all'interno delle relazioni statistiche elaborate. La relazione del Segretario Generale delle Nazioni Unite nel corso della 19<sup>a</sup> sessione del 1976, accolta ed approvata dalla Commissione, ha convenuto come lo sviluppo del sistema di statistica dell'energia debba avere priorità nel lavoro della commissione. Ha quindi evidenziato come sia necessario passare attraverso l'utilizzo dei bilanci energetici da parte degli Stati, per ottenere un coordinamento dei lavori sulle statistiche dell'energia e la fornitura di dati in un forma adatta per la comprensione e l'analisi del ruolo dell'energia nell'economia.

Partendo da queste considerazioni, si è quindi ritenuto necessario l'adozione di regole internazionali per l'elaborazione delle statistiche energetiche che possano armonizzare le attività che ogni Stato ha fino a quel momento posto in essere.

Conseguentemente la Divisione Statistica delle Nazioni Unite (UNSD) ha predisposto un report dettagliata su concetti e metodi rilevanti per le statistiche energetiche di base. Tale lavoro è stato apprezzato nel corso della 20<sup>a</sup> riunione della Commissione nel 1979, decidendo di rendere il lavoro disponibile agli uffici statistici nazionali ed internazionali, nonché ad altre agenzie competenti.

Successivamente nel 1987, la Commissione ha ritenuto necessario l'elaborazione di un manuale sui fattori di conversione e sulle unità di misura da utilizzare nell'elaborazione di statistiche energetiche.

Alle due decisioni sopra esposte il UNSD ha quindi dato corso con la pubblicazione nel 1982 del *Concepts and methods in energy statistics, with special reference to energy accounts and balances: a technical report*, Studies in Methods, Series F, No. 29, United Nations, New York, 1982 [16]; e nel 1987 con la pubblicazione di *Energy statistics: definitions, units of measure and conversion factors*, Studies in Methods, Series F, No. 44, United Nations, New York, 1987 [17].



Ai due documenti sopra esposti si deve dare il merito di aver svolto un importante ruolo nello sviluppo delle statistiche energetiche a livello nazionale ed internazionale, dando indicazioni primarie complete ed organiche e che costituiscono ancora oggi la base dei sistemi statici sull'energia dei singoli stati.

Negli anni successivi, ulteriori indicazioni sono state necessarie da parte del UNSD. Infatti, l'esperienza maturata da diversi Stati sulle statistiche energetiche e le indagini condotte su settori e dati specifici, hanno reso le precedenti raccomandazioni carenti su alcuni aspetti. Così, nel 1991 il UNSD ha pubblicato *Energy statistics: a manual for developing countries*, Studies in Methods, Series F, No. 56, United Nations, New York, 1991 [18].

Nel 2004 è invece la volta dell'Agenzia internazionale dell'Energia (IEA) ed Eurostat – U.E., che hanno elaborato e pubblicato i loro manuali sulle statistiche energetiche: *Energy Statistics Manual*, OECD/IEA/EUROSTAT, Paris, 2004 [19]. Tuttavia, le statistiche sull'energia mantenevano gravi criticità in termini di disponibilità dei dati e di loro comparazione a livello internazionale. Al fine di superare le ancora presenti criticità, nel 2005 la Commissione Europea ha riconosciuto la necessità dell'integrazione delle statistiche energetiche come parte delle statistiche ufficiali e la revisione delle raccomandazioni fino ad allora fornite. Sono stati quindi istituiti due gruppi di lavoro che lavorano rispettivamente: il primo sullo sviluppo dei metodi e regole internazionali per le statistiche nazionali sull'energia; il secondo, sul tema dell'armonizzazione delle definizioni relative ai prodotti energetici.

## **4.2 L'Europa e le statistiche sull'energia**

Al fine di monitorare l'impatto e le conseguenze della sua politica in campo energetico, la Comunità europea ha riconosciuto la necessità di possedere dati precisi e tempestivi sulle quantità di energia, sulle sue forme, sulle sue fonti, sulla sua generazione, sul suo approvvigionamento, sulla sua trasformazione e sui suoi consumi. Fino agli anni 2000 le statistiche dell'energia si sono tradizionalmente incentrate sull'approvvigionamento energetico e sulle energie fossili, ma con il rapido aumento dell'energia prodotta da Fonti energetiche rinnovabili (FER) che si è registrata a seguito delle decisioni prese dalla stessa Comunità in merito alla politica energetica che dovranno perseguire gli Stati membri, a partire dal 2000 si è evidenziato come occorreva incentrarsi maggiormente su una conoscenza e un monitoraggio accresciuti del consumo energetico finale, delle energie rinnovabili e dell'energia nucleare.

Oltretutto, la disponibilità di informazioni accurate e aggiornate sull'energia è essenziale ai fini della valutazione dell'impatto dei consumi energetici sull'ambiente, in particolare con riferimento alle emissioni di gas ad effetto serra. Tali informazioni sono infatti richieste dalla decisione dell'11 febbraio 2004 n. 280/2004/ CE del Parlamento europeo e del Consiglio, relativa ad un meccanismo per monitorare le emissioni di gas a effetto serra nella Comunità e per attuare il protocollo di Kyoto.

I primi provvedimenti che impongono agli stati membri di trasmettere dati quantitativi sull'energia al fine di controllare i progressi in direzione del conseguimento degli obiettivi stabiliti, sono: la direttiva 2001/77/CE del Parlamento europeo e del Consiglio, del 27 settembre 2001, sulla promozione dell'energia elettrica prodotta da fonti energetiche rinnovabili nel mercato interno dell'elettricità [20], e la direttiva 2004/8/CE del Parlamento europeo e del Consiglio, dell'11 febbraio 2004, sulla promozione della cogenerazione basata su una domanda di calore utile nel mercato interno dell'energia [21].

Inoltre, le successive decisioni sono: la direttiva 2002/91/CE del Parlamento europeo e del Consiglio, del 16 dicembre 2002, sul rendimento energetico nell'edilizia [22]; la direttiva 2006/32/CE del Parlamento europeo e del Consiglio, del 5 aprile 2006, concernente l'efficienza degli usi finali dell'energia e i servizi energetici [23]; e la direttiva 2005/32/CE del Parlamento europeo e del Consiglio, del 6 luglio 2005, concenenti: l'istituzione di un quadro per l'elaborazione di specifiche per la progettazione ecocompatibile dei prodotti che consumano energia; imposizione agli Stati membri di trasmettere dati quantitativi sul consumo energetico; e la necessità di una migliore interfaccia tra i dati sull'energia e le indagini statistiche pertinenti, come il censimento della popolazione e degli alloggi e i dati sui trasporti [24].

La necessità di disporre di statistiche sull'energia dell'Unione europea è altresì resa necessaria dall'intento di istituire un osservatorio europeo del mercato energetico al fine di monitorare l'adempimento dei programmi sanciti con il libro verde della Commissione del 22 giugno 2005 sull'efficienza energetica ed il libro verde della Commissione dell'8 marzo 2006 su una strategia europea per un'energia sostenibile, competitiva e sicura. In più, la necessità di istituire un modello di previsione energetica nel settore pubblico - chiesta dal Parlamento europeo nella risoluzione del 14 dicembre 2006 su una strategia europea per un'energia sostenibile, competitiva e sicura - richiede senza alcun dubbio la raccolta di dati dettagliati e aggiornati sull'energia.

La necessità di modelli predittivi, per cui sono necessari dati tempestivi è sollecitata anche dalla maggiore attenzione che la sicurezza dell'approvvigionamento dei carburanti più importanti sta assumendo, e dall'esigenza di poter anticipare e coordinare le soluzioni che l'Unione europea può intraprendere in eventuali crisi di approvvigionamento.

A tali necessità nel 2008 il Parlamento ed il Consiglio dell'Unione hanno adottato il regolamento n. 1099/2008 relativo alle statistiche dell'energia, che istituisce un quadro comune per la produzione, la trasmissione, la valutazione e la diffusione di statistiche dell'energia comparabili nella Comunità [25]. Il regolamento sancisce che il sistema di statistiche dell'energia deve offrire garanzie di comparabilità, trasparenza, flessibilità e capacità di evolvere per essere utile al processo decisionale politico dell'Unione europea e dei suoi Stati membri, nonché per promuovere un dibattito pubblico che coinvolga i cittadini. Per questo motivo è pienamente riscontrata la necessità di integrare dati statistici

relativi all'energia nucleare e sviluppare ulteriormente dati pertinenti relativi alle energie rinnovabili. Analogamente, in materia di efficienza energetica, la disponibilità di dati statistici dettagliati sull'habitat e sul trasporto è ritenuta di grande utilità.

### **4.3 Le statistiche dell'energia in Italia**

Il tema delle statistiche energetiche in Italia è di giovane trattazione, basti pensare che solo nel 2010 l'Istituto nazionale di statistica (Istat) ha deliberato l'adozione del Codice italiano delle statistiche ufficiali. Le statistiche energetiche assumono una loro organica e strutturale organizzazione solo a partire dall'emanazione del Regolamento n. 1099/2008 da parte del Parlamento e del Consiglio dell'Unione europea. Fino ad allora i dati e le statistiche energetiche erano rese note dagli operatori del settore che operavano in determinati *asset* del sistema energetico, e le metodologie di reperimento ed elaborazione erano le più svariate. Tuttavia, il Ministero dello Sviluppo Economico (MiSE), negli anni ha svolto un'azione di raccolta ed elaborazione di tali dati al fine, anche, di fornire annualmente un report ufficiale sulla situazione energetica nazionale.

Con l'emanazione del suddetto Regolamento il MiSE ha assunto anche la veste di unico soggetto deputato a fornire all'Unione europea i dati statistici riguardanti lo Stato italiano, secondo le modalità stabilite dallo stesso regolamento. A livello regionale, le singole regioni autonomamente realizzano annualmente un monitoraggio sulla loro situazione energetica, tuttavia si deve constatare come tali stime sono, in alcuni casi, la mera estrapolazione del dato regionale fornita dallo stesso MiSE. È quindi facile desumere che il dato energetico regionale è posseduto dagli operatori che intervengono nella sfera dell'energia ed in particolare nelle macro aree: della produzione, del trasporto/dispacciamento, e della vendita/consumo.

La prima area vede soggetti volti alla produzione di energia elettrica, gas e combustibili liquidi. I dati annuali sono quindi la mera proposizione delle quantità di fonte primaria estratta o lavorata e del prodotto immesso nel sistema. Si noti come in tale area molto incerti sono i dati riferiti alle fonti rinnovabili (FER), elettriche e termiche, poiché il soggetto produttore può essere anche un utente domestico che possiede limitate conoscenze dei parametri statistici. Tale problematica per le FER elettriche sono superate attraverso la gestione del dato aggregato da parte di soggetti terzi, quali Terna S.p.A. e G.S.E. S.p.A., che per la propria *mission* aziendale sono anche possessori di dati di produzione delle fonti rinnovabili elettriche. Difficoltà invece rilevanti si presentano ancora oggi per la contabilizzazione, e quindi l'elaborazione, di dati riferiti alle FER termiche, questo poiché non c'è un effettivo sistema di monitoraggio della produzione di calore da fonti rinnovabili e pertanto il dato è calcolato, e non puntuale, attraverso elaborazioni che tengono conto delle caratteristiche tecnologiche, delle ore di funzionamento e della disponibilità della fonte rinnovabile.

I dati relativi al trasporto e dispacciamento dei vettori energetici (energia elettrica, gas e combustibili liquidi) sono in possesso di soggetti proprietari delle reti di trasmissione, quali: Terna S.p.A. per l'energia elettrica e SNAM S.p.A. per il gas. I combustibili liquidi e i rimanenti combustibili gassosi, sono invece contabilizzati dalle Agenzie delle Dogane territoriali.

Dati sugli usi finali sono invece in possesso di differenti Società che operano nella vendita dei prodotti energetici. Dati aggregati su tale macro area vengono forniti dall'Autorità per l'energia elettrica, il gas ed il servizio idrico, che ha il compito di vigilare sulla tutela del consumatore e sulla regolazione che si attua a tale livello.

Come detto, il MiSE fornisce i dati ufficiali sul sistema energetico nazionale e sui vettori energetici che lo interessano, tuttavia è stato più volte evidenziato come le stime provenienti da diversi soggetti istituzionali non siano armonizzate. Tale motivo è dovuto essenzialmente ad un differente metodo di acquisizione del dato, che alle volte risulta basato su rilevazioni tramite questionari per un dato campione ed alle volte è puntuale su tutto il campione coinvolto.

Al fine di omogenizzare i dati e di creare una metodologia unica e condivisa, il MiSE con d.M. del 14 gennaio 2012 ha approvato la metodologia che, nell'ambito del sistema statistico nazionale in materia di energia, è applicata per rilevare i dati necessari a misurare il grado di raggiungimento degli obiettivi nazionali in materia di quote dei consumi finali lordi di elettricità, energia per il riscaldamento e il raffreddamento, e per i trasporti coperti da fonti energetiche rinnovabili [26]. Successivamente, in attuazione all'articolo 40, comma 5, del decreto legislativo n. 28 del 2011 e nel rispetto delle finalità di cui al medesimo articolo 40, commi 1 e 2, è stata approvata con d.M. del 11 maggio 2015 la metodologia di monitoraggio degli obiettivi regionali di Burden Sharing. Si è disposto altresì che tale metodologia è applicata, nell'ambito del Sistan in materia di energia [27].

Con tale decreto il legislatore nazionale ha voluto dare indicazioni ben chiare su:

- Dato da monitorare ai fini della verifica del burden sharing;
- Soggetto responsabile dell'elaborazione del dato;
- Periodicità dell'elaborazione dei dati;
- Descrizione metodologia;
- Metodo di acquisizione dei dati;
- Strumenti di controllo dell'informazione di input.

Il decreto ha quindi stabilito che la responsabilità dei dati riferiti alle fonti fossili di energia è di E.N.E.A., mentre tutto ciò che concerne le fonti rinnovabili di energia è responsabilità di G.S.E. S.p.A.. Si rileva quindi come per il settore dell'energia elettrica il sistema di rilevazione e monitoraggio statistico è ormai sviluppato e consolidato e consente di rispondere adeguatamente ai

flussi informativi a livello nazionale e a livello regionale richiesti dalla citata direttiva 2009/28/CE. Mentre per i settori termico e dei trasporti le informazioni statistiche non sono disponibili col medesimo grado di approfondimento e pertanto si rende necessaria un'implementazione del sistema, nonché lo sviluppo di specifiche metodologie di calcolo. Ai fini di un'armonizzazione tra i dati statistici regionali e quelli nazionali, il MiSE, sempre con il d.M. del 11 maggio 2015, ha stabilito come il G.S.E. e l'E.N.E.A. debbano applicare la metodologia regionale assicurando coerenza tra i risultati complessivi così ottenuti e quelli derivanti dall'applicazione della metodologia nazionale. La coerenza fra la sommatoria dei dati relativi alle Regioni e alle Province autonome e i corrispondenti valori nazionali è verificata annualmente per i singoli componenti degli obiettivi oggetto di rilevazione e monitoraggio e per l'ammontare complessivo. In caso di mancata coerenza, gli esiti sono trasmessi al Ministero dello sviluppo economico per le eventuali azioni correttive. Tuttavia, qualora le Regioni o le Province autonome dispongano di statistiche sui componenti dei loro consumi energetici, prodotte conformemente ai requisiti di qualità delle statistiche Sistan e rese disponibili nei tempi e secondo le modalità stabilite dal d.M., e tali statistiche regionali differiscono di almeno dieci punti percentuali rispetto al corrispondente valore stimato con la metodologia regionale, il GSE ovvero l'E.N.E.A. comunicano la circostanza al Ministero dello sviluppo economico che procede a verificare l'attendibilità dei dati forniti dalle Regioni o Province autonome, che se confermata ne autorizza il loro utilizzo.

## **CAPITOLO 5** *I dati statistici sull'energia*

Al fine di meglio comprendere la natura dei dati trattati in questo lavoro di ricerca e le unità di misura di riferimento, nel corso della trattazione di questo capitolo verranno fornite informazioni complete sulle diverse tipologie di dati energetici.

### **5.1 Fonte primaria e secondari di energia**

Le numerose fonti energetiche esistenti possono essere classificate in diversi modi. Si dicono primarie se sono utilizzabili direttamente, così come si trovano in natura. Sono fonti primarie: il carbone, il petrolio, il gas naturale, la legna, i combustibili nucleari (uranio), il Sole, il vento, le maree, i laghi montani e i fiumi (da cui è possibile ottenere energia idroelettrica) e il calore della Terra che fornisce energia geotermica. Sono secondarie quelle che derivano dalla trasformazione delle fonti primarie di energia, ad esempio: la benzina, che deriva dal trattamento del petrolio grezzo e l'energia elettrica ottenuta dalla conversione di energia meccanica (centrali idroelettriche, eoliche) o chimica (centrali termoelettriche) o nucleare (centrali nucleari).

### **5.2 Fonti rinnovabili e non rinnovabili**

Alcune fonti sono rinnovabili, cioè forniscono energia che si rigenera in continuazione mediante trasformazioni chimiche (come le biomasse) o fisiche (come l'energia idrica, solare, eolica, etc). In particolare il Sole, il vento, l'acqua, le maree, il calore della Terra sono fonti inesauribili, sempre disponibili e che non finiranno mai. Le biomasse, invece, sono in grado di rigenerarsi in tempi confrontabili con quelli della vita dell'uomo. Nel caso della legna, per esempio, è possibile avere sempre a disposizione del combustibile, pur di consumarne solo una quantità limitata e di preoccuparsi di riforestare laddove sono stati abbattuti gli alberi. Le cosiddette fonti non rinnovabili, invece, hanno tempi di rigenerazione talmente lunghi (milioni di anni) che una volta sfruttate si considerano esaurite. Sono quelle che si sono formate nel corso di milioni di anni, come i combustibili fossili (petrolio, carbone, gas naturale) o addirittura al momento della formazione del nostro pianeta, come l'uranio. La disponibilità di queste fonti, per quanto grande, è limitata ed esse costituiscono una sorta di magazzino energetico della Terra.

### 5.3 Le unità di misura dell'energia

Le unità di misura utilizzate dall'uomo per esprimere le quantità delle fonti energetiche sono numerose. Vi sono misure per le quantità fisiche e misure per il contenuto di energia o di calore.

Tra le più note misure per le quantità fisiche possiamo citare la tonnellata, utilizzata per il petrolio greggio e il carbone; il barile, utilizzato per il petrolio greggio; il metro cubo, utilizzato per il gas; il litro, per la benzina ed il gasolio. Confrontare le fonti energetiche tra loro utilizzando le diverse unità di misura per le quantità fisiche è un esercizio estremamente complesso, se non impossibile: da un chilogrammo di petrolio, infatti, non si ottiene la stessa energia prodotta da un chilogrammo di carbone o da un metro cubo di gas. Volendo paragonare le diverse fonti energetiche è meglio utilizzare un'unità di misura comune, ovvero esprimerle in termini di "contenuto" di energia o di calore. L'unità di misura ufficiale dell'energia è il Joule (J). Le più diffuse unità di misura per l'energia sono il chilowatt all'ora (chilowattora, kWh) utilizzato soprattutto per l'energia elettrica; per il calore vi sono la BTU (British Thermal Unit), la chilocaloria (o grande caloria = kcal) e soprattutto la TEP o Tonnellata Equivalente di Petrolio. Quest'ultima unità di misura, la TEP, è la più diffusa a livello internazionale poiché è legata ad uno dei più importanti e utilizzati combustibili: il petrolio. Misurando le diverse fonti di energia in termini di TEP è possibile confrontarle fra loro e aggregarle, operazione fondamentale per calcolare quanta energia utilizza uno Stato in un anno o quanta energia ha ancora a disposizione nel sottosuolo (giacimenti di petrolio, gas e carbone). In sostanza una TEP rappresenta la quantità di calore ottenibile da una tonnellata di petrolio.

La trasformazione delle diverse fonti di energia in calore è effettuata partendo dai poteri calorifici inferiori, scelta che si ritiene più corretta di quella relativa ai poteri calorifici superiori nei quali sono comprese le calorie di condensazione del vapore acqueo che si forma durante la combustione, e cioè quella quantità di energia estraibile sotto forma di calore da una unità fisica del combustibile considerato.

Con riferimento ai coefficienti di conversione ufficialmente adottati si hanno:

| Combustibili solidi                         | tep/tonn |
|---|----------|
| Combustibili Solidi non specificati         | 0,71667  |
| Antracite e carboni magri (Anthracite)      | 0,71654  |
| Carbone bituminoso (Bituminous Coal)        | 0,66638  |
| Carbone subbituminoso (Sub-bituminous Coal) | 0,47220  |
| Carbone da cokeria (Coking Coal)            | 0,77704  |
| Agglomerati (Patent fuel)                   | 0,69265  |
| Coke da cokeria (Coke oven Coke)            | 0,69265  |
| Catrame greggio (Coal Tar)                  | 0,90000  |
| Lignite (Brown Coal)                        | 0,25002  |
| Torba (Peat)                                | 0,64000  |

Tabella 5.1 - Fattore di conversione in tep dei combustibili solidi - Fonte: E.N.E.A,

| Combustibili liquidi             | tep/tonn   |
|----------------------------------|------------|
| Gas di raffineria (Refinery gas) | 1,18228718 |
| Gas di petrolio liquefatto       | 1,09869112 |
| Virgin nafta                     | 1,05092194 |
| Benzine                          | 1,05092194 |
| Carboturbo tipo benzina          | 1,05092194 |
| Ragia minerale                   | 1,04136810 |
| Esano                            | 1,18228718 |
| Carboturbo tipo petrolio         | 1,02703735 |
| Petrolio da riscaldamento        | 1,02703735 |
| Gasoli                           | 1,01748352 |
| Olio combustibile                | 0,95538359 |

Tabella 5.2 - Fattore di conversione in tep dei combustibili liquidi - Fonte: E.N.E.A.

| Combustibili Gassosi          | tep/10 <sup>3</sup> mc |
|-------------------------------|------------------------|
| Gas naturale                  | 0,81900                |
| Gas naturale liquefatto (LGN) | 0,81900                |
| Gas di estrazione             | 0,54239                |
| gas d'officina (Workin gas)   | 0,42500                |

Tabella 5.3 - Fattore di conversione in tep dei combustibili gassosi - Fonte: E.N.E.A.

È bene evidenziare che i valori assunti dal p.c.i. delle varie fonti sono estremamente variabili, e sono da considerare come valori medi standardizzati ed uniformemente accettati, in quanto le caratteristiche fisico-chimiche di ogni tipologia di carburante, anche la semplice differenza di provenienza geografica può far abbassare od innalzare il p.c.i., come, ad esempio, per l'olio combustibile usato nelle centrali, il diverso contenuto di zolfo, oltre ad incidere sulle emissioni di inquinanti, varia il potere calorifico.

Discorso per certi versi analogo va impostato per l'energia elettrica, in quanto, considerando un'efficienza media delle centrali di trasformazione circa del 39%, il kWh ha una resa calorica all'utenza finale di 860 kcal, ma per produrlo servono invece circa 2.200 kcal (valore medio delle centrali termiche tradizionali). Nella redazione di un Bilancio Energetico l'energia elettrica negli usi finali è quindi contabilizzata a 860 kcal/kWh che è uguale a 0.0860 tep/MWh, in modo da tenere conto delle perdite di trasformazione.

## 5.4 I dati economici correlati all'energia

I dati economici possono costituire una delle chiavi di lettura del dato energetico assoluto. Infatti, come detto nelle premesse di questo lavoro, l'energia ha refluienze sui sistemi economici di ogni singolo stato. Diventa allora fondamentale comparare il dato puramente energetico ad indicatori



economici che possono quanto più rappresentare la solidità finanziarie e del sistema produttivo di uno stato, regione, provincia o comune. A tal proposito l'indicatore *Valore aggiunto* è quello che più si presta alle analisi condotte nel proseguo del lavoro. Infatti, rappresenta la differenza fra il valore della produzione di beni e servizi e i costi sostenuti da parte delle singole unità produttive per l'acquisto di input produttivi, a essa necessari, presso altre aziende. Esso rappresenta quindi il valore che i fattori produttivi utilizzati dall'impresa, capitale e lavoro, hanno 'aggiunto' agli input acquistati dall'esterno, in modo da ottenere una data produzione (Economic Value Added, EVA).

Il *Valore aggiunto* così definito è detto "ai prezzi di mercato". Se nel calcolo del valore aggiunto, dal valore della produzione si sottraggono l'IVA (Imposta sul Valore Aggiunto) e le imposte indirette sulle importazioni, si ottiene il valore aggiunto ai prezzi del produttore. Sottraendo da questo le altre imposte indirette, e sommando i contributi alla produzione, si ottiene il valore aggiunto al costo dei fattori. Il valore aggiunto corrisponde anche alla remunerazione dei fattori produttivi ed è quindi composto dai redditi da lavoro e dai redditi da capitale-impresa, ovvero il risultato lordo di gestione.

Nel caso delle attività della pubblica amministrazione, la produzione non viene valorizzata direttamente sul mercato, e per questa ragione la misura del valore aggiunto è ottenuta direttamente sommando le remunerazioni dei fattori produttivi, essenzialmente il lavoro, data la bassa incidenza dei profitti nelle attività pubbliche. Il *Valore aggiunto* è quindi una misura appropriata per quantificare l'esito dell'attività da parte di una singola impresa o settore, rispetto alla nozione di produzione, che incorpora al suo interno il valore aggiunto di una pluralità di settori o imprese. Il valore aggiunto misura quindi l'apporto che ciascuna unità economica fornisce alla formazione del PIL. Rispetto al PIL calcolato in base al valore delle transazioni di beni finali, il concetto di valore aggiunto consente di esplicitare il concorso delle singole unità produttrici di prodotti intermedi alla formazione di quello finale e svolge un ruolo essenziale nella rappresentazione dell'attività economica relativa ai singoli settori di produzione. Infatti, ai fini analitici, in riferimento all'intera economia la nozione di valore aggiunto e quella di PIL sono praticamente coincidenti. Viceversa, non è possibile calcolare il PIL di singoli settori, visto che un prodotto finale deriva nella maggior parte dei casi dal concorso dell'attività di unità produttive appartenenti a settori diversi. Il processo produttivo può pertanto essere interpretato come una sommatoria di valore aggiunto forniti da singole imprese sino a ottenere i prodotti finali, di consumo o investimento o destinati all'esportazione. Quando un processo produttivo è segmentato secondo il valore aggiunto di molte imprese, viene definito 'disintegrato', mentre se il valore della produzione finale deriva dalla somma dei valore aggiunto di un numero limitato di unità produttive, il processo di produzione risulta integrato verticalmente. Come detto, il valore aggiunto può essere espresso ai prezzi correnti oppure a prezzi costanti. Ciò avviene attraverso l'utilizzo di appositi indici di prezzi (deflatori del valore aggiunto). I deflatori del valore aggiunto di un dato settore differiscono dall'andamento dei prezzi dei prodotti del settore (in quanto da essi va scorporata la parte che remunera gli acquisti dei prodotti intermedi) e riflettono pertanto solamente l'andamento dei redditi

unitari di quel settore. Il valore aggiunto a prezzi costanti diviso per il monte ore lavorate misura la produttività del lavoro, ovvero l'output per ora lavorata. Il rapporto fra i redditi da lavoro e il valore aggiunto a prezzi costanti calcola invece l'incidenza del costo del lavoro su ciascuna unità prodotta, ovvero il Costo del Lavoro per Unità di Prodotto (CLUP). La ripartizione del valore aggiunto a prezzi correnti nelle componenti dei redditi da lavoro e del risultato lordo di gestione descrive la distribuzione del reddito fra i fattori della produzione. Processi produttivi di tipo capital intensive tendono a destinare una quota maggiore del valore aggiunto alla remunerazione del capitale, e viceversa per i processi labour intensive.

## **5.5 I dati climatici**

Importanza primaria nelle analisi energetiche è tradizionalmente posta sui dati relativi alle condizioni atmosferiche del luogo di studio. Tali dati assumo un maggiore peso se riferiti ad un ben preciso luogo circoscritto nei confini e di estensione limitata, ad esempio un comune. Per analisi a più ampio raggio il dato, sebbene importante, assume un'influenza sempre minore per via che i punti di misura non sono infiniti e che le stime sono un valore medio delle condizioni atmosferiche rispetto a superfici molto vaste.

I dati climatici in Italia sono caratterizzati principalmente da serie storiche limitate e da numerosi momenti in cui il dato non è stato registrato per non funzionamento della centraline meteo. Oltre a ciò si registra come i dati relativi ad ambienti urbani sono quasi del tutto mancanti e pertanto l'interazione delle condizioni meteo delle città, con una densità elevata di utenti finali, ed il consumo proprio del distretto metropolitano non può essere determinato con la dovuta accuratezza, dovendosi basare su una misura avvenuta in ambiente periferico o addirittura rurale. I dati più comuni acquisiti dalle centraline meteo-climatiche sono: la temperatura max., min. media; l'umidità media; i mm di precipitazione; la velocità del vento media e max. e la sua direzione; la pressione atmosferica e la radiazione solare.

Uno degli indici climatici più utilizzati per le analisi energetiche, soprattutto per i consumi per riscaldamento è il grado giorno. Tale indice, riferito ad una ben precisa località, indica: la somma, estesa a tutti i giorni di un periodo annuale convenzionale di riscaldamento, delle sole differenze positive giornaliere tra la temperatura dell'ambiente, fissata convenzionalmente per ogni nazione, e la temperatura media esterna giornaliera. In Italia la temperatura ambiente è fissata a 20 °C, negli Stati Uniti a 18,3 °C oppure 15,5 °C, mentre in Gran Bretagna a 18,5 °C, 15,5 °C e 10 °C [28].

## **5.6 I dati demografici**

Mentre per lo studio dei consumi riferibili alle attività produttive l'indicatore economico risulta quello che più di tutti può essere utilizzato per comprendere il consumo energetico nei settori che lo caratterizzano, per il settore residenziale l'indicatore più appropriato è quello relativo alle indagini demografiche. È infatti ormai comprovato come nel complesso ogni individuo di un determinato distretto tenda a consumare come ogni individuo che abita lo stesso distretto, pertanto all'interno dell'area di studio. È anche vero che le caratteristiche dell'edificio in cui si vive e le potenzialità economiche del nucleo familiare determinano un maggiore o una minore spesa energetica del singolo soggetto. Tuttavia ai fini statistici l'assumere un valore medio non comporta errori rilevanti nell'ambito residenziale. I dati pertanto più utilizzati nelle comparazioni con i dati energetici risultano: il numero della popolazione e delle famiglie; il numero i componenti il nucleo familiare; il numero delle abitazioni primarie e secondarie.

Grazie alle indagini annuali fatte da Istat, le serie storiche di tale tipologia di dati sono ben complete ed esaustive e rappresentano un dato utilizzabile e completo anche per eventuali interpolazioni dovute alla mancanza di altri indicatori.

## **5.7 Gli indicatori energetici ed ambientali**

Per una analisi delle caratteristiche del sistema energetico di un determinato territorio che consenta di evidenziare e spiegare, con riferimento ad un certo periodo di tempo, i cambiamenti che si sono manifestati nelle modalità di impiego delle fonti energetiche nei diversi settori d'uso (Agricoltura, Industria, Trasporti, Terziario e Residenziale), si possono utilizzare i dati assoluti forniti dai bilanci energetici, ma anche specifici rapporti che siano espressione sintetica dell'andamento di serie storiche energetiche rispetto ad altre (economiche, demografiche, fisiche, etc.). L'analisi energetica si avvarrà dei parametri relativi alle attività economiche (Agricoltura, Industria e Terziario) e civili (Residenziale e Servizi) poiché l'energia è uno dei fattori produttivi utilizzati nelle stesse attività e pertanto è importante analizzare le modalità del suo impiego, ovvero il rendimento o efficienza. I rapporti tra grandezze energetiche e variabili economiche, strutturali, demografiche, etc., calcolabili, in base ai dati disponibili, per ogni settore di attività economica ai diversi livelli di aggregazione territoriale, vengono definiti *Indicatori di efficienza energetica*.

Il concetto di efficienza energetica è utilizzato per indicare i progressi compiuti negli interventi atti ad assicurare un uso "razionale" dell'energia. Progressi che si identificano in una migliore utilizzazione dell'energia sia nella prestazione di un servizio che nella produzione di un bene e che consentono una riduzione dei consumi energetici a parità di servizio fornito o di produzione realizzata.

Lo scopo del calcolo degli indicatori è di poter fare un'analisi dettagliata e comparata con altre realtà territoriali. Le analisi consentono quindi di evidenziare situazioni particolari per le quali si possano avanzare le prime ipotesi di interventi tesi all'ottimizzazione degli impieghi energetici nei diversi settori di consumo e quindi di effettuare congetture sull'evoluzione futura. Il calcolo degli indicatori energetici richiede la disponibilità di serie storiche omogenee ed attendibili relative sia alle variabili energetiche che a quelle economico, strutturali e demografiche. A questo scopo possono essere utilizzate statistiche ufficiali o, in mancanza, stime ad hoc. Gli indicatori calcolabili a livello nazionale sono di due tipi: descrittivi ed esplicativi; i primi rappresentano la situazione energetica del territorio, gli altri hanno lo scopo di spiegare le ragioni delle variazioni dei consumi energetici e degli indicatori descrittivi e, conseguentemente, del miglioramento o peggioramento del sistema.

A livello regionale la carenza di dati disaggregati non permette di determinare gli stessi indicatori calcolati a livello nazionale ed internazionale (in particolare quelli esplicativi). È comunque possibile effettuare confronti temporali e spaziali tra le varie regioni e tra queste e l'Italia. Gli indicatori possono essere calcolati rispetto alle variabili economiche e rispetto a quelle tecnico-economiche. I primi esprimono la cosiddetta intensità energetica, ossia il rapporto tra consumo di energia e un indice di attività economica (PIL, Valore Aggiunto, Consumi Privati, Investimenti, etc.), generalmente valutato a prezzi costanti per evitare gli effetti dell'inflazione. Invece i rapporti tecnico economici sono calcolati rapportando il consumo di energia ad indicatori di attività misurati in termini fisici (tonnellate di acciaio, numero di passeggeri, etc.), oppure a delle unità di consumo (veicoli, abitazioni, etc.) e prendono il nome di consumi unitari.

### ***5.7.1 Fattore di intensità energetica***

Il consumo di energia di uno Stato o area geografica è strettamente legato al suo livello di attività economica. Il rapporto tra le due grandezze - Prodotto Interno lordo (PIL) o numero della popolazione (nel caso di settore residenziale) e Consumo Interno Lordo di energia (CIL) è definito intensità energetica ed è un indicatore, seppur grezzo e aggregato, dell'efficienza energetica di una economia. Trattandosi di un rapporto, tanto più basso è il valore dell'intensità energetica tanto più aumenta l'efficienza energetica della economia interessata.

L'intensità energetica è un indicatore fortemente correlato allo sviluppo economico di un paese e l'efficienza energetica non è l'unica variabile che influenza l'intensità: altri fattori, quali i cambiamenti nella struttura economica ed industriale, gli stili di vita e le condizioni climatiche, hanno un impatto sull'andamento di questo indicatore. Quindi una riduzione dell'intensità energetica non sempre e non solo indica un miglioramento dell'efficienza energetica: ad esempio il passaggio da una struttura produttiva ad alto consumo di energia verso una a basso consumo di energia, a parità di

condizioni, determina una riduzione dell'intensità energetica che non è associata ad un incremento dell'efficienza energetica.

La struttura economica di un Paese o di un'area gioca, quindi, un ruolo importante nel determinarne la relativa intensità energetica. Si può affermare che le economie post industriali, il cui settore terziario ha un forte peso sul PIL, avranno in linea di principio una intensità energetica considerevolmente più contenuta delle economie caratterizzate da attività tradizionali, con ampio peso dell'industria pesante.

Molto importante è allora introdurre i concetti di intensità energetica finale, definita dal rapporto tra il consumo finale complessivo e il PIL, e intensità energetica settoriale, definita dal rapporto tra il consumo finale e il Valore Aggiunto del settore, ad eccezione dei settori Trasporti e Residenziale. L'intensità energetica del settore Trasporti è determinata dal rapporto tra consumo finale e il PIL; l'intensità energetica del settore Residenziale è data dal rapporto tra il consumo finale e la spesa delle famiglie o il numero della popolazione.

### ***5.7.2 Gli indicatori ambientali***

Per indicatori di sostenibilità ambientale e indicatori ambientali si intendono dati e parametri utili alla valutazione qualitativa o quantitativa delle condizioni ambientali e socio-economici di un sistema. In particolare gli indicatori ambientali sono parametri rilevabili che mostrano le condizioni di un sistema ambientale; gli indicatori di sostenibilità sono indicatori ambientali particolari che aiutano a comprendere se le condizioni ambientali rientrano in determinate aspettative e soddisfano gli obiettivi di sviluppo sostenibile. Grazie agli indicatori di sostenibilità ambientale è possibile individuare eventuali problemi ed ipotizzare soluzioni; ad esempio utilizzando i dati sulle emissioni globali di gas serra si possono ipotizzare scenari futuri del sistema di interazione uomo-clima e favorire la sostenibilità delle politiche di consumo degli idrocarburi. Ogni indicatore ambientale può essere considerato come una variabile significativa del sistema da comprendere; di conseguenza tanto più è complesso un sistema, tanti più sono gli indicatori necessari per descriverlo; gli indicatori di sostenibilità ambientale sono associati agli obiettivi di sviluppo sostenibile.

L'uso degli indicatori di sostenibilità ambientale è sempre più praticato da enti, organizzazioni ed istituzioni, a livello locale o globale (ONU, OCSE, UE, Comuni, aziende, etc.) che intendono intervenire con logiche di sviluppo sostenibili; molte di tali istituzioni pubblicano periodicamente rapporti ambientali basati su indicatori di sostenibilità ambientale, al fine di comprendere e controllare i legami tra economia, società ed ambiente. Differenti sono i modi per definire gli indicatori, e dunque anche quelli di sostenibilità. Una prima grande distinzione è tra indicatori relativi a fenomeni direttamente misurabili e indicatori relativi a fenomeni non misurabili in maniera diretta. Molti dei fenomeni che riguardano la sostenibilità sono quantificabili. Alcuni sono fenomeni chimici e fisici direttamente misurabili (ad es. le emissioni di CO<sub>2</sub>); altri invece sono caratteri per i quali non

disponiamo di strumenti di misura diretta, ma che possono sempre essere espressi quantitativamente con riferimento ad un'appropriata e ponderata scala di intensità. Più interessante, ai fini delle valutazioni di sostenibilità, è la distinzione tra:

- indicatori fisici, dotati di unità di misura e che esprimono i livelli delle variabili individuate come significative;
- indicatori multidimensionali, o indici, costituiti da aggregazione di indicatori e dati dello stesso tipo o di tipi diversi.

In concreto, i primi forniscono l'informazione di base sulle componenti del sistema, ad esempio sulle quantità e sui flussi (es. produzione annua di rifiuti in una regione); mentre i secondi consentono di condensare l'informazione presente in più parametri per meglio comunicare e facilitare la conoscenza (es. accoppiamento tra la produzione di rifiuti e benessere economico misurato dal rapporto tra i rifiuti prodotti ed il prodotto interno lordo di una regione). Gli indici mettono in luce ad esempio i rapporti tra le componenti del sistema; inoltre gli indici possono essere espressi in valori assoluti, attraverso la normalizzazione e l'aggregazione delle informazioni di partenza, facilitano le comparazioni tra sistemi ambientali, territori ed organizzazioni. Gli indicatori possono essere classificati anche in base alla funzione che svolgono:

- gli indicatori descrittivi, sono indicatori elementari che misurano “cosa sta succedendo” in relazione alle varie componenti ambientali, sono i tipici indicatori di base per la caratterizzazione della situazione ambientale;
- gli indicatori prestazionali, sono indicatori descrittivi associati a dei target, delle soglie, dei valori obiettivo misurabili; quindi gli indicatori prestazionali misurano la distanza (“distance-to-target”) di un sistema ambientale rispetto a valori di riferimento, obiettivi politici, livelli di sostenibilità; perciò gli indicatori prestazionali spesso sono anche chiamati “di efficacia” e sono molto utili perché spingono all'azione, consentono di finalizzare le politiche, di focalizzare gli interventi di sviluppo, di misurare i progressi verso lo sviluppo sostenibile; naturalmente la rilevanza degli indicatori prestazionali è funzione degli obiettivi che ci si pone, cioè gli obiettivi ambientali e gli indicatori prestazionali si corrispondono in una struttura gerarchica;
- gli indicatori (indici) di efficienza, sono espressi come indici e misurano l'efficienza di uso delle risorse o di inquinamento per unità di prodotto, di processo, di reddito, ecc.

Ai fini della valutazione dello sviluppo sostenibile è più utile la suddivisione degli indicatori in due categorie:

- Indicatori assoluti, che esprimono una misura assoluta di un indicatore ritenuto rilevante nel corso dell'indagine.

- Indicatori relativi, intesi come relazioni tra indicatori assoluti dello stesso tipo o di tipo diverso.

Gli indicatori assoluti forniscono le informazioni di base del sistema. Gli indicatori relativi, mettendo in relazione tali informazioni, consentono di avere una conoscenza completa del sistema. Confrontando gli indicatori assoluti, infatti, è possibile studiare i legami tra i parametri, il loro andamento nel tempo e la loro velocità di variazione.

Gli indicatori possono essere ulteriormente organizzati secondo un modello di causalità uomo-ambiente, sviluppato negli anni '70 dallo statistico canadese Anthony Friend: il modello pressione-stato-risposte (PSR). Secondo questo modello, prima le attività umane esercitano pressioni sull'ambiente, cambiandone lo stato, poi rispondono per adattare il sistema uomo-ambiente. Negli anni '90 il modello PSR venne adottato dall'OCSE e poi venne modificato dalla commissione per lo sviluppo sostenibile delle Nazioni Unite (Uncsd) che sostituì le pressioni con i determinanti (driving forces), cioè i fattori socio-economici che causano le pressioni ambientali. I determinanti sono in pratica i fattori di sviluppo socio-economico significativi per i loro risvolti ambientali. Nella seconda metà degli anni '90 l'Agenzia ambientale europea finalmente adottò uno schema di classificazione più completo: determinanti, pressioni, stato, impatti, risposte ambientali (DPSIR; Eea, 1999; Cagnoli 2010). Lo schema DPSIR è oggi uno schema molto noto, utilizzato per classificare gli indicatori.

Gli indicatori “determinanti” descrivono le attività socio-economiche che causano le pressioni ambientali; indicatori determinanti sono ad esempio il numero di abitanti presenti in un bacino, oppure l'estensione fisica di un insediamento.

Gli indicatori di “pressione” descrivono le azioni dell'uomo che direttamente causano modifiche sullo stato delle componenti ambientali (cioè direttamente impattanti per l'ambiente), come i prelievi di risorse naturali o le emissioni di inquinanti; indicatori di pressione sono ad esempio la portata dei prelievi d'acqua da un fiume, oppure di reflui scaricati nel fiume. Il rapporto tra un indicatore di pressione con un indicatore determinante, tra di loro correlati, fornisce un indice d'efficienza ambientale (p.e. per bacino idrografico il rapporto tra gli scarichi di reflui e gli abitanti presenti indica se gli insediamenti hanno un'efficienza depurativa più o meno alta).

Gli indicatori di “stato” descrivono le condizioni di qualità delle varie componenti ambientali; indicatori di stato sono ad esempio la portata di un fiume o la concentrazione dei nitrati che inquinano l'acqua di un fiume.

Gli indicatori di “impatto” descrivono le modifiche di stato per effetto delle pressioni antropiche; indicatori di impatto sono ad esempio la riduzione di portata fluviale a valle di un prelievo oppure l'aumento di nitrati nelle acque di un fiume a valle di uno scarico. Si rileva in particolare che gli indicatori d'impatto dovrebbero essere l'obiettivo conoscitivo fondamentale di ogni studio di impatto ambientale. Il rapporto tra un indicatore di impatto ed un indicatore di pressione, tra di loro correlati,

fornisce un indice di sensibilità ambientale della componente interessata (p.e. in un fiume il rapporto tra i nitrati presenti e quelli scaricati indica una sensibilità ambientale più o meno elevata).

Gli indicatori di “risposta” descrivono le azioni umane intraprese per risolvere un problema ambientale, come ad esempio la depurazione degli inquinanti o la riduzione dei consumi di risorse naturali; indicatori di risposta sono ad esempio la portata dei rilasci ecologici d’acqua da un invaso artificiale oppure la quantità di nitrati depurati dai reflui di fognatura.



## CAPITOLO 6 *Analisi di dati temporali*

Strumenti adeguanti ed importanti per una pianificazione efficiente ed efficace risultano utili in molte situazioni che interessano la ricerca, l'impresa e il management, dove è necessario predire comportamenti e pianificare le azioni future. Il vantaggio di avvalersi di tali strumenti è quello di essere meno soggetti ad eventi inattesi in quanto si attua un approccio più scientifico riguardo alla conoscenza dell'ambiente in cui si opera. Sono state sviluppate numerose tecniche predittive che sono classificate in due categorie generali: tecniche quantitative, basati su metodi statistici; e tecniche qualitativi, basate prevalentemente su giudizi.

I metodi di tipo quantitativo possono essere impiegati quando:

- sia disponibile una sufficiente informazione sull'evoluzione passata del fenomeno;
- tale informazione possa essere quantificata;
- si possa assumere che le caratteristiche dell'evoluzione passata continuino a sussistere nel futuro, al fine di effettuare la previsione.

L'applicazione del corretto metodo di previsione si basa sulla migliore identificazione della relazione che c'è fra la variabile da prevedere e il tempo (oppure altre variabili che hanno il ruolo di predittori).

Le tecniche o metodi di tipo quantitativo sono numerosi ed applicati nell'ambito di numerose discipline. Nella maggioranza dei casi le tecniche quantitative sono di tipo formalizzato; esse richiedono cioè l'impiego di metodologia statistico-matematica a livello più o meno elevato [29].

È importante soffermarci sulla categorizzazione interna ai metodi quantitativi fra: metodi di analisi delle serie storiche e metodi esplicativi. Questi ultimi assumono che la variabile da prevedere possa essere messa in relazione con una o più variabili indipendenti o esplicative. Tali tecniche di previsione impiegano i metodi di regressione e quindi la fase principale dell'analisi consiste nella specificazione e stima di un modello che mette in relazione la variabile da prevedere (variabile risposta) e le variabili esplicative. Queste, spesso, hanno la funzione di variabili strumentali rispetto alle quali si esplicita l'andamento del fenomeno da prevedere. Nella fase di previsione vera e propria è quindi richiesta la conoscenza dei livelli futuri delle variabili esplicative che, sostituiti nella formula stimata del modello, forniscono la previsione per la variabile di interesse dello studio. I metodi esplicativi risultano di particolare utilità ed efficacia quando la variabile esplicativa è direttamente manipolabile dal decisore. Nell'analisi delle serie storiche, invece, il fenomeno da prevedere viene trattato come una scatola nera in quanto non si cerca di individuare i fenomeni che lo possano influenzare. L'obiettivo di questo

approccio consiste nell'identificazione dell'evoluzione passata del fenomeno e nella estrapolazione del sentiero percorso per ottenere una previsione. In altre parole, il fenomeno da prevedere viene modellato rispetto al tempo e non rispetto ad una variabile esplicativa. Questa impostazione si rivela senz'altro utile quando:

- il fenomeno è poco conosciuto, oppure è difficile individuare (misurare) le relazioni che questo ha con altri fenomeni;
- si è interessati a conoscere ciò che accadrà e non il modo in cui accadrà.

Gli approcci o metodi di tipo qualitativo non richiedono dati o informazioni formalizzate. Essi si basano soprattutto su giudizi e sull'esperienza accumulata e sono prevalentemente usati per indicare tendenze più che per prevedere specifici valori numerici. Sono quindi a disposizione degli utenti un vasto numero di strumenti di previsione che variano per: l'informazione necessaria, il livello di formalizzazione e di trattamento statistico-matematico, l'orizzonte temporale di previsione, il costo.

Nel proseguo porremo l'attenzione sulle fasi previste nell'analisi di dati temporali e sui metodi più utilizzati per l'analisi predittive e la loro validazione.

## **6.1 Le serie storica e dati cross-section**

Prima di iniziare la descrizione delle fasi relative alle analisi predittive, è importante soffermarci su alcune definizioni dei dati oggetto di uno studio di questo tipo. Una distinzione certamente da fare è tra serie storica e dati cross-section. Una serie storica è una sequenza di osservazioni ordinate rispetto al tempo (ad esempio: il fatturato mensile, i prezzi giornalieri delle azioni, il tasso di interesse settimanale, il profitto annuo, etc.). Lo scopo dell'analisi delle serie storiche consiste nello studio dell'evoluzione passata del fenomeno rispetto al tempo; la previsione viene ottenuta ipotizzando che tali regolarità di comportamento si ripetano nel futuro. Si assume, quindi, che i tempi di osservazione siano equidistanti. Ciò è supportato anche dal fatto che molti fenomeni di interesse aziendale vengono registrati in corrispondenza di intervalli di medesima ampiezza.

Con il termine cross-section si intende un tipo di studio basato su un campionamento trasversale. Rappresenta uno studio condotto in un determinato tempo, prendendo una porzione di popolazione di dati (una sezione incrociata). La loro analisi fornisce solo indirettamente un'evidenza circa gli effetti di tempo e devono essere usati con grande cautela quando si devono trarre conclusioni circa il cambiamento.

Introduciamo le seguenti definizioni relative ai dati raccolti:

- *Popolazione statistica*: insieme degli elementi a cui si riferisce l'indagine statistica;

- *Unità statistica*: ogni elemento della popolazione statistica, la minima unità della quale si raccolgono i dati;
- *Campione statistico*: un qualsiasi insieme di unità statistiche prese da tutta la popolazione. Un campione è dunque un sottoinsieme di misurazioni selezionate dalla popolazione.
- *Variabile quantitativa*: quando assume valori numerici. Può essere *Continua* quando assume valori continui in un intervallo, o *Discreta* quando assume valori discreti come numero di campioni, numero di geni sovra-espresso, numero di pazienti, etc.
- *Variabile qualitativa*: quando assume valori non numerici. Può essere Ordinale quando i dati sono in un ordine, o Categorical.

## **6.2 Le fasi di lavoro della predizione**

Un'analisi predittiva basata su dati quantitativi si sviluppa in cinque fasi.

1. *Definizione del problema*: si tratta dell'attività più complessa in quanto richiede la comprensione delle modo in cui la previsioni verranno utilizzate, chi ne farà uso e come il o i modelli saranno utilizzati;
2. *Raccolta di informazioni*: è necessario tener conto sia delle informazioni statistiche che dell'esperienza accumulata da chi ha conoscenza della sorgente dati. Le informazioni storiche sono utilizzate per l'implementazione del modello di previsione ma a queste vanno accompagnate tutte quelle utili alla comprensione delle caratteristiche funzionali di ciò che si sta analizzando;
3. *Analisi preliminare dei dati*: in questa fase vengono analizzati i dati raccolti calcolando le più importanti statistiche descrittive relative a ciascun set di dati. Utili strumenti in questa fase sono lo scatter plot e la scomposizione per individuare elementi quali trend, ciclo, stagionalità, elementi "irregolari";
4. *Scelta e adattamento del modello*: comporta un'attività di confronto tra più modelli anche se è bene precisare che un'attenta analisi preliminare è in grado di indirizzare la scelta verso un numero limitato;
5. *Valutazione del modello e suo impiego a fini predittivi*: selezionato il modello e stimati i parametri, il modello stesso può essere utilizzato per la previsione. Una prima indicazione sulla bontà del modello proviene da opportuni indicatori, mentre l'effettiva applicazione dello stesso con il riscontro su dati reali darà la certezza della bontà del lavoro condotto.

### 6.3 Analisi grafiche preliminari

Elemento preliminare nello studio di serie temporali, è quello di visualizzare i dati mediante una rappresentazione grafica. Lo scopo è quello di individuare eventuali regolarità di comportamento che sono utili nel suggerire l'approccio modellistico. Il grafico più semplice è il cosiddetto time plot (oppure line plot) che consiste nella rappresentazione dei dati rispetto al tempo (Figura 6.1).

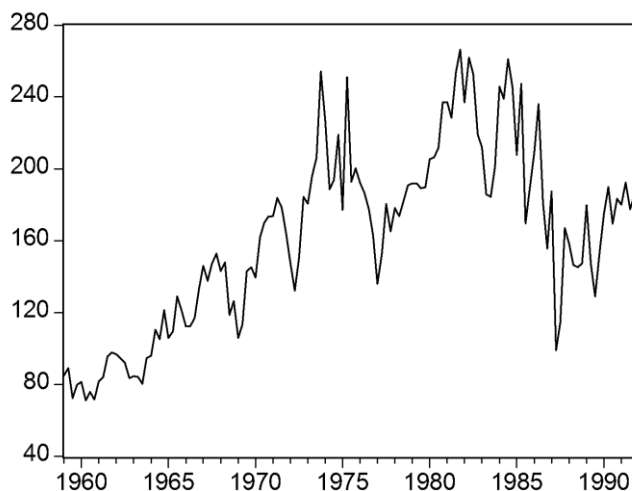


Figura 6.1 - Time plot di serie storiche di dati.

Come è evidente, un time plot immediatamente rivela tendenze o oscillazioni regolari, e altri andamenti di tipo sistematico rispetto al tempo. In generale, possono essere individuati quattro tipi di andamento rispetto al tempo:

1. *Trend orizzontale*: in questo caso la serie oscilla intorno ad un valore costante (media della serie). Tale serie è detta stazionaria in media;
2. *Trend stagionale*: questo esiste quando la serie è influenzata da fattori stagionali (es. mensile, semestrale, trimestrale, etc.). Le serie influenzate dalla stagionalità sono dette anche serie periodiche poiché il ciclo stagionale si ripete in un periodo fisso;
3. *Trend ciclico*: questo tipo di andamento è presente quando la serie presenta aumenti e diminuzioni che non sono di periodo fisso. Questa è la principale differenza fra le fluttuazioni cicliche e quelle stagionali. Inoltre, l'ampiezza delle oscillazioni cicliche è generalmente più grande di quella dovuta alla stagionalità;
4. *Trend di fondo o residuale*: è caratterizzato da un andamento crescente o decrescente di lungo periodo. La serie della popolazione residente in Italia è un esempio di andamento tendenziale o trend di tipo crescente.

Molte serie evidenziano una combinazione di più Trend, è proprio questo genere di complessità che rende l'operazione di previsione estremamente complessa ed interessante. I metodi predittivi, infatti, devono essere in grado di riconoscere le varie componenti della serie in modo da riprodurle nel futuro, nell'ipotesi che il pattern passato continui a ripetersi, nelle sue caratteristiche evolutive, anche nel futuro. Se una serie di dati esibisce un'oscillazione stagionale è utile eseguire un plot stagionale, che consiste nella rappresentazione dei valori della serie versus i periodi stagionali come: mesi, trimestri, semestri, etc..

## 6.4 Analisi descrittive dei dati

L'analisi descrittiva dei dati in studio è quanto più utile, necessaria e di corredo all'analisi tramite plot. Gli indici sintetici più utilizzati per condurre questo tipo di analisi, sono:

- *Media*: rappresenta la somma di tutti i valori delle variabili della popolazione di dati diviso il numero di unità della popolazione di dati

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t ; \quad [1]$$

- *Moda*: è il valore più frequente di una distribuzione, o meglio, la modalità più ricorrente della variabile;
- *Mediana*: è il valore che occupa la posizione centrale in un insieme ordinato di dati. È una misura robusta, in quanto poco influenzata dalla presenza di dati anomali;
- *Campo di variazione*: rappresenta la differenza tra il dato più grande e quello più piccolo della distribuzione

$$R = \text{Max}(y_i) - \text{Min}(y_i); \quad [2]$$

- *Varianza*: caratterizza molto bene la variabilità di una popolazione

$$S^2 = \frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y})^2 ; \quad [3]$$

- *Deviazione standard*: è una misura di distanza dalla media misura della dispersione della variabile casuale intorno alla media

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y})^2} ; \quad [4]$$

Altre misure statistiche impiegate nell'analisi delle serie storiche sono: la covarianza e la correlazione che, nella fattispecie, vengono definite rispettivamente:

$$\text{autocovarianza } c_k = \frac{1}{n-1} \sum_{t=1}^{n-k} (y_t - \bar{y}) \cdot (y_{t+k} - \bar{y}); \quad [5]$$

$$\text{autocorrelazione } r_k = \frac{\frac{1}{n-1} \sum_{t=1}^{n-k} (y_t - \bar{y}) \cdot (y_{t+k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} = \frac{c_k}{c_0}; \quad [6]$$

essendo calcolate fra coppie di punti della medesima serie, distanti  $k$  periodi. La costante  $k$  è detta *Time lag*.

L'autocovarianza definisce in particolare il grado di dipendenza o memoria delle variabili che compongono una serie di dati.

L'autocorrelazione invece misura la correlazione incrociata di una determinata variabile con sé stessa in diversi punti nel tempo. Può essere definita come la somiglianza tra le osservazioni di una determinata variabile in funzione del ritardo temporale. Fornisce un valore assoluto, indipendente dalla scala delle variabili che compongono il processo. Un'autocorrelazione uguale a più o meno 1 equivale a una perfetta relazione lineare positiva o negativa.

## **6.5 Trasformazioni e aggiustamenti**

Talvolta l'aggiustamento o la trasformazione dei valori della serie originale produce dati più facilmente interpretabili.

Tre tipologie di *aggiustamenti* sono le seguenti:

- trasformazioni matematiche;
- aggiustamenti per rimuovere le oscillazioni dovute a effetti di calendario;
- aggiustamenti per rimuovere oscillazioni dovute a mutamenti nella popolazione oppure nei prezzi.

Se operiamo la trasformazione dei dati, il metodo di previsione lavorerà sui valori trasformati.

## **6.6 Metodi di previsione**

Uno degli scopi dell'analisi delle serie storiche è fornire indicazioni predittive sul fenomeno d'interesse. Per previsione si intende una descrizione di avvenimenti futuri che si fonda su un insieme coordinato di ipotesi. La previsione si fonda sul presupposto logico che, in talune situazioni, è

ragionevole supporre che le informazioni disponibili per il passato, adeguatamente trattate, siano in grado di ridurre l'incertezza circa gli accadimenti futuri.

La previsione si distingue in relazione a tre elementi:

- a) orizzonte temporale:
  - breve o congiunturale (12-18 mesi);
  - medio (5 anni);
  - lungo termine (non oltre 10 anni).
  
- b) obiettivo
  - strumentale (unico scopo avvertimento);
  - tendenziale (evitare il ripetersi di fenomeni);
  - condizionale (ipotesi forti che condizionano le previsioni);
  - normativa (come è possibile conseguire certi obiettivi).
  
- c) metodo impiegato
  - qualitativo
    - informale o previsione naïf;
  - quantitativo
    - serie storiche (scomposizione e liscio esponenziale);
    - regressioni.

### ***6.6.1 Metodi quantitativi e qualitativi***

La molteplicità degli scenari per i quali si è chiamati ad effettuare attività di previsione si differenziano tra loro, per esempio, con riferimento all'orizzonte temporale, al risultato effettivo e ovviamente, in base alle caratteristiche dei dati disponibili. La necessità di conseguire i migliori risultati possibili tenendo conto dei vincoli reali porta sostanzialmente ad individuare due grandi tipologie di previsioni e quindi metodi predittivi: quantitative e qualitative. Rientrano nella prima categoria quelle previsioni in cui l'informazione riguardante il passato è di fatto disponibile; può essere rappresentata nella forma di dati numerici e si presume che alcune caratteristiche della struttura dei dati possono ravvisarsi anche nei periodi successivi a quello di osservazione. Tale ultima condizione costituisce un elemento fondamentale di tutte le tecniche quantitative (e anche di molte qualitative) indipendentemente dal grado di complessità del modello utilizzato. Le tecniche

quantitative di previsione costituiscono un insieme decisamente variegato, essendo state sviluppate in seno a discipline profondamente diverse. Ciascuna presenta caratteristiche ben individuabili in termini di accuratezza, costi e proprietà statistiche. In questa categoria è inoltre possibile distinguere tra modelli intuitivi e modelli basati invece su precise assunzioni statistiche. Nel primo caso ciò che conta è l'esperienza empirica che dipende dal settore di studio considerato e addirittura da chi svolge la previsione. L'elevato grado di soggettività che accomuna questo tipo di previsione ne ha grandemente ridotto il loro utilizzo nel corso del tempo anche se ad oggi numerosi utenti continuano ad effettuare previsioni sulla base di tale criterio sia in quanto non sono a conoscenza anche dei più semplici metodi statistici sia in quanto si ravvisa una naturale preferenza nei confronti di criteri fondati sul giudizio anziché su basi oggettive. È bene tuttavia segnalare che lo scetticismo derivante dal considerare il passato inadeguato nel prevedere il futuro dovrebbe essere in qualche misura mitigato dalla considerazione che alcuni elementi relativi all'andamento passato di un determinato fenomeno tendono a ripetersi nel tempo con un certo grado di stabilità. Volendo invece tentare di effettuare un'ulteriore distinzione nell'ambito dei metodi quantitativi, è possibile considerare due grandi classi di metodi: quelli che tendono a considerare una relazione funzionale tra la variabile da prevedere ed una o più variabili indipendenti e quelli in cui il fenomeno sotto osservazione può essere convenientemente indagato solo per il tramite delle osservazioni e degli errori passati con l'obiettivo di catturarne l'andamento caratteristico. La possibilità di prevedere un determinato fenomeno mediante la sola indagine dei valori passati può essere dettata dall'estrema complessità del sistema a cui la variabile afferisce o anche, nel caso in cui siano chiare le relazioni funzionali tra le diverse grandezze, dalla complessità dei legami tra le variabili.

Sia i metodi basati sull'andamento della serie storica che quelli fondati su relazioni esplicative presentano particolari vantaggi a seconda del loro utilizzo. In conclusione, l'attività di previsione può oggi contare su un'ampia varietà di modelli che differiscono per filosofia concettuale, dai metodi utilizzati, capacità e possibilità di utilizzo. Tale varietà è dovuta anche alla possibilità attuale di utilizzare strumenti informatici dotati di elevate potenze di calcolo a costi accettabili.

### **6.6.2 Approccio classico e moderno**

Un modello stocastico generale per descrivere il processo generatore dei dati di una serie storica  $y_1, y_2, \dots, y_n$  relativa ad una variabile  $y$  è dato da:

$$y_t = f(t) + u_t \quad [7]$$

Si assume che la serie osservata sia il risultato della composizione di:



- una sequenza completamente deterministica,  $\{f(t)\}$ , che costituisce la parte sistematica della serie;
- una sequenza di variabili casuali  $\{u_t\}$ , che rappresenta la parte stocastica della serie ed obbedisce ad una determinata legge di probabilità.

Le due sequenze non sono individualmente osservabili, ma vanno determinate sulla base del campione. Se trattiamo il modello stocastico generale secondo l'approccio classico alle serie storiche, si può supporre che esista una legge di evoluzione temporale del fenomeno, rappresentata da  $f(t)$ . La componente casuale  $u_t$  viene assunta a rappresentare l'insieme delle circostanze, ciascuna di entità trascurabile, che non si vogliono o non possiamo considerare in  $y_t$ . I residui di  $y_t$  vengono pertanto imputati al caso ed assimilati ad errori accidentali. Questo equivale ad ipotizzare che la componente stocastica  $u_t$  sia generata da un processo *white noise*, ovvero da una successione di variabili casuali indipendenti, identicamente distribuite, di media nulla e varianza costante.

In sintesi, nell'approccio classico l'attenzione viene concentrata su  $f(t)$ , con  $u_t$  considerato un processo a componenti incorrelate e dunque trascurabile. Nell'approccio moderno si ipotizza invece che  $f(t)$  manchi o sia già stata 'eliminata' (mediante stima o altri metodi). L'attenzione viene posta sulla componente stocastica  $u_t$ , che si ipotizza essere un processo a componenti correlate del tipo

$$u_t = g(y_{t-1}, y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) + \varepsilon_t \quad [8]$$

che va trattato con opportune tecniche statistiche.

### 6.6.3 Metodi informali - qualitativi

Nei metodi c.d. qualitativi gli input richiesti dipendono dal metodo utilizzato e in modo sostanziale da attività legate al giudizio dell'analista e alla conoscenza accumulata. Tale impostazione è dettata principalmente dalla necessità di lavorare con uno scarso livello di informazioni quantitative ma possedendo un sufficiente grado di conoscenza del fenomeno oggetto di indagine rispetto ai suoi elementi qualitativi. Rientrano in questa categoria un'ampia gamma di metodologie che si differenziano per tempo di analisi, complessità e capacità di previsione. Molte organizzazioni utilizzano contemporaneamente uno o più modelli di questo tipo e, dove possibile, vengono affiancati da modelli quantitativi. Valutare l'effettiva utilità di modelli qualitativi è compito arduo ed essi costituiscono piuttosto un supporto all'attività di pianificazione e ai modelli quantitativi. La struttura e le caratteristiche di costo che accompagnano tali modelli ne hanno confinato il loro utilizzo nell'ambito delle scelte di medio/lungo periodo quali la formulazione di strategia, lo sviluppo di nuove tecnologie e prodotti, la messa a punto di piani di medio e lungo termine. La qualità dei risultati previsti ottenibili dall'utilizzo di modelli qualitativi è stata molte volte messa in discussione ma

rimane un dato di fatto che molti decisori utilizzano tali modelli in combinazione con tecniche quantitative.

#### **6.6.4 Metodo di scomposizione**

La possibilità di tener conto degli andamenti caratteristici di una serie storica separando la componente casuale da ciò che ne rappresenta invece l'andamento caratteristico, riveste un ruolo significativo nell'ambito dei metodi di previsione. Secondo questa visione, le componenti caratteristiche della serie possono essere proiettate nel futuro ed utilizzate come previsione. Un tale approccio, detto di scomposizione, può essere particolarmente utile soprattutto per quanto riguarda la conoscenza del comportamento effettivo della serie e ai fini di una migliore accuratezza della previsione.

La scomposizione, che risale agli inizi del XX secolo, rappresenta quindi uno degli approcci più consolidati nell'analisi delle serie storiche. Due sono i filoni lungo i quali tale metodologia si è sviluppata: la prima linea di ricerca imponeva l'eliminazione di qualsiasi correlazione spuria dovuta alla presenza di trend nei dati, al fine di studiare correttamente la correlazione seriale di una serie o quella esistente tra più serie; la seconda area di ricerca si basa sulle analisi svolte dagli economisti il cui fine era quello di valutare gli impatti della depressione economica e, come ovvia conseguenza, tentare di prevederne l'andamento futuro.

Dallo sviluppo fino ad ora raggiunto, per quanto attiene particolari settori, i metodi di scomposizione generalmente si concentrano su due fondamentali componenti: il c.d. ciclo-trend e la stagionalità. La prima componente identifica le variazioni del livello della serie che intercorrono in un arco temporale lungo, mentre i fattori stagionali riguardano tutte quelle fluttuazioni periodiche, con periodo costante, dovute all'influenza sulla serie di determinati elementi quali ad esempio la temperatura, la frequenza e il livello delle precipitazioni, il periodo dell'anno e la particolare politica adottata dall'impresa. Il ciclo-trend può essere a sua volta scomposto nei singoli elementi anche se molte delle procedure esistenti tendono a considerare il ciclo-trend come un'unica componente.

In termini sintetici, la scomposizione può essere espressa mediante la seguente rappresentazione:

$$y = x + \varepsilon = f(\text{ciclo-trend}, \text{stagionalità}, \text{errore}) \quad [9]$$

dove  $\varepsilon$  rappresenta la componente “irregolare” o “residuale”.

La letteratura propone numerosi approcci per identificare le componenti della serie, ma il concetto di fondo risiede su considerazioni di carattere empirico e pertanto implica inizialmente l'identificazione del ciclo-trend e successivamente della componente stagionale. La componente residuale è pertanto assunta di tipo casuale e quindi se pur non prevedibile, almeno è identificabile. È opportuno precisare che tale modo di procedere non è particolarmente robusto dal punto di vista teorico e che tuttavia esso ha trovato, anche con risultati soddisfacenti, largo uso nella pratica.

Volendo dettagliare i componenti che rappresentano la funzione scomposizione, abbiamo:

- *Trend*: è la tendenza di fondo del fenomeno considerato, riferita ad un lungo periodo di tempo;
- *Ciclo*: è costituito dalle fluttuazioni attribuibili al succedersi nel fenomeno considerato di fasi ascendenti e di fasi discendenti, generalmente collegate con le fasi di espansione e di contrazione dell'intero sistema economico;
- *Stagionalità*: è costituita dai movimenti del fenomeno nel corso dell'anno che, per effetto dell'influenza di fattori climatici e sociali, tendono a ripetersi in maniera pressoché analoga nel medesimo periodo (mese o trimestre). (p.e.: PIL italiano, nel mese di Agosto tutte le grandi fabbriche sono chiuse);
- *Componente accidentale*: come nel modello di regressione, anche nei modelli di serie storiche non vi è mai una relazione perfetta tra la variabile sotto osservazione e le diverse componenti, la componente accidentale tiene conto di questo e del comportamento non perfettamente prevedibile degli agenti economici;

La composizione a sua volta, può essere di tipo: additiva, moltiplicativa e mista.

La funzione con modello additivo, con ipotesi di indipendenza tra le componenti, può essere descritta dalla seguente funzione:

$$Z_t = T_t + C_t + S_t + a_t \quad [10]$$

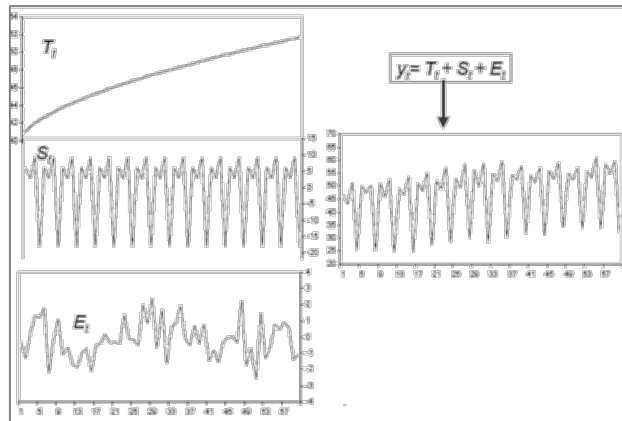


Figura 6.2 - Esempio di serie trimestrale a componenti additivi.

Il modello moltiplicativo con componenti dipendenti è rappresentato dalla funzione:

$$Z_t = T_t \times C_t \times S_t \times a_t \quad [11]$$

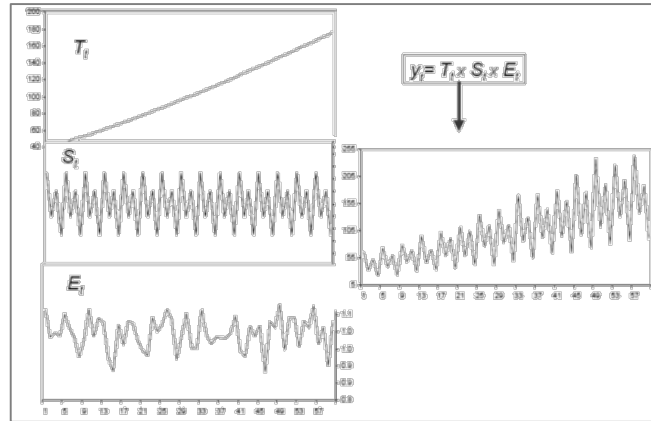


Figura 6.3 - Esempio di una serie trimestrale a componenti moltiplicativi.

Il modello misto è il seguente:

$$Z_t = T_t \times S_t + C_t \times a_t \quad [12]$$

I pregi di un approccio con modello di scomposizione sono: semplicità, buoni risultati con serie anche corte, prima approssimazione. I difetti, invece, stanno nella pluralità di soluzioni ed una visione troppo settorizzata.

Le ultime ricerche su tale metodo, sia per quanto riguarda la ricerca teorica che quella applicata, hanno prodotto numerosi tentativi di basare la previsione direttamente sulla scomposizione. In sostanza, l'approccio è quello di proiettare le singole componenti in avanti nel tempo per poi passare ad una loro successiva ricomposizione. Sebbene una tale impostazione sembri ragionevole tuttavia un tale approccio raramente è in grado di fornire risultati soddisfacenti essendo la principale difficoltà costituita proprio dalla previsione delle singole componenti. Tra le varie componenti, il ciclo-trend è senz'altro quella che presenta le maggiori difficoltà e le proposte di modellare tale componente mediante una funzione parametrica sono apparse piuttosto deboli. Con riferimento alla componente stagionale è necessario considerare i suoi eventuali cambiamenti nel corso del tempo. La componente irregolare viene spesso assunta come una successione di valori serialmente incorrelati, caratteristica raramente riscontrata nella pratica.

### 6.6.5 Metodi di liscio esponenziali

La maggior parte dei fenomeni oggetto di indagine, descrivibili mediante una successione temporale di valori numerici, presenta una serie di caratteristiche legate alla struttura del fenomeno oggetto di osservazione. La necessità di tener conto di aspetti quali trend, ciclo e stagionalità ha portato così a sviluppare una serie di tecniche di fatto riconducibili al concetto di media.

È possibile suddividere le previsioni che utilizzano il concetto di media in due gradi raggruppamenti: quelli che di fatto si riferiscono propriamente al concetto di media mediante l'applicazione di pesi uguali per ciascun set di osservazioni e quelli che invece utilizzano un insieme di pesi che non si mantiene costante ma che decresce in modo esponenziale al fine di conferire “minor importanza” alle osservazioni più distanti nel tempo. Rientrano pertanto in questa seconda categoria i criteri di lisciamento esponenziale (exponential smoothing) che impongono alcune restrizioni per quanto riguarda i pesi da applicare ai dati storici. Seguendo la classificazione offerta in Makridakis (1998) i metodi di lisciamento esponenziale possono prevedere la stima di un solo parametro oppure tener conto dei cambiamenti registrati nell'andamento dei dati. Vi è poi il metodo Holt nel quale vengono utilizzati due differenti parametri e permette di effettuare la previsione in presenza di trend.

Per quanto detto, il lisciamento esponenziale risulta il metodo più utilizzato specialmente in ambito aziendale per le previsioni nel breve periodo. Qualità: flessibilità e semplicità d'uso.

Data una serie  $\{y_t\}_{t=1}^n$  supposto di volere prevedere  $y_{n+k}$  con  $k>1$  orizzonte temporale. Si indica con  $\hat{y}_{t+k}$  la previsione fatta al tempo  $t$ .

Se la serie ha un trend costante alterato da fattori accidentali  $y_t = a + \varepsilon_t$  [13]

la previsione è data dalla perequazione di tutte le informazioni disponibili fino a  $n$  con uguale peso:

$$\hat{y}_{n+1} = \frac{1}{n} \sum_{t=1}^n y_t = \frac{1}{n} \sum_{j=1}^n y_{n-j+1} \quad [14]$$

Se invece la serie cambia in modo stocastico è più realistico calcolare la previsione tramite una media ponderata dando peso maggiore alle osservazioni più recenti:

$$\hat{y}_{n+1} = \frac{\sum_{l=1}^n w_l y_{n-l+1}}{\sum_{l=1}^n w_l} \quad [15]$$

con  $w_l \geq 0$ ;  $w_l \leq w_{l-1}$   $l=2,3,\dots,n$

scritta anche

$$\hat{y}_{n+1} = \sum_{j=0}^{n-1} c_j y_{n-j} \quad c_j = \frac{w_{j+1}}{\sum_{l=1}^n w_l} \quad [16]$$

$j = 0,1,\dots,n-1$

Il lisciamiento esponenziale è detto tale perché la serie  $y_t$  viene sostituita dalla successione  $\hat{y}_{t+1}$  ottenuta da:

$$\hat{y} = c y + c y + c y + \dots \quad [17]$$

Con pesi definiti dalla successione esponenziale

$$c = c\delta, \quad j = 0, 1, \dots, n-1, \quad 0 < \delta < 1 \quad [18]$$

La previsione con costante di lisciamiento  $\delta$  è pari a:

$$\hat{y}_{n+1} = (1-\delta) \sum_{j=0}^{n-1} \delta^j y_{n-j} \quad [19]$$

Dove la previsione è una media ponderata di tutte le osservazioni disponibili; queste influenzano la previsione con intensità decrescente all'aumentare della distanza dal tempo  $n$ . Più la costante è vicina ad 1 più la previsione è rigida, cioè più influenzata dalle osservazioni passate; viceversa più è prossima a 0 più la previsione è flessibile, cioè influenzata dalle osservazioni più recenti.

La precedente si può anche scrivere:

$$\hat{y}_{n+1} = \delta \hat{y}_n + (1-\delta)y_n \quad [20]$$

Cioè come una media ponderata tra la previsione fatta al tempo  $n-1$  e l'ultima osservazione  $y_n$ , il cui peso è tanto più forte quanto più piccola è la costante. Appare chiara la logica di aggiornamento sequenziale del metodo (la previsione viene modificata dall'osservazione più recente) ed il ruolo della costante di lisciamiento. Il lisciamiento esponenziale rappresenta, quindi, una regola di aggiornamento della previsione basata sul valore previsto al tempo precedente, corretto con un termine proporzionale all'errore di previsione compiuto. La previsione è ottenuta come media ponderata di tutte le osservazioni disponibili. Il parametro di lisciamiento ( $\delta$ ) darà:

- per  $\delta \rightarrow 0$  il lisciamiento esponenziale attribuisce sempre più peso ai nuovi dati e l'effetto perequativo è quasi nullo, considerando dunque la serie affidabile e la previsione non fa altro che restituire l'ultima osservazione disponibile;
- per  $\delta \rightarrow 1$  il lisciamiento esponenziale attribuisce peso pressoché nullo ai nuovi dati e la nuova previsione tende a coincidere con la precedente.

Per stimare l'equazione [20] è necessario avere un valore iniziale di  $\delta$  e di  $\hat{y}_1$ .

$\delta$  viene posto tra 0,05 e 0,3 oppure stimato mediante algoritmi non lineari. Ad esempio con i minimi quadrati si cerca quel valore che minimizza:

$$\sum_{t=m}^{n-1} (y_{t+1} - \hat{y}_{t+1})^2 \quad [21]$$

Per quanto riguarda  $\hat{y}_1$  usualmente:

$$\begin{aligned} \hat{y}_1 &= y_1 \text{ se } \delta \text{ è molto piccolo} \\ \hat{y}_1 &= (y_1 + y_2) / 2 \\ \hat{y}_1 &= \frac{1}{n} \sum_{t=1}^n y_t \text{ se } \delta \text{ è più vicino ad } 1 \end{aligned} \quad [22]$$

Se  $n$  è grande e/o la costante di lisciamiento è piccola, la scelta di tale valore è ininfluenta.

La previsione può essere vista come la costante che meglio si adatta alla serie in prossimità di  $n$ . Questo fa capire che se la serie ha un trend non costante o fluttuazioni marcate, tale metodo non è appropriato.

Una generalizzazione del metodo di lisciamiento applicando la perequazione di  $y_t$  tramite la determinazione del livello, del trend e della stagionalità esponenziale è detto metodo di Holt-Winters, con le varie varianti di seguito riportate:

➤ *Metodo non stagionale*

In prossimità di  $n$  un aggiustamento invece di applicarlo con una perequazione si attua mediante una retta di equazione

$$L_n + T_n(t - n) = 0 \quad [23]$$

con:

$$\begin{aligned} \hat{L}_n &= \alpha(\hat{L}_{n-1} + \hat{T}_{n-1}) + (1 - \alpha)y_n \quad 0 < \alpha < 1 \\ \hat{T}_n &= \beta\hat{T}_{n-1} + (1 - \beta)(\hat{L}_n - \hat{L}_{n-1}) \quad 0 < \beta < 1 \end{aligned} \quad [24]$$

Tale metodo è più flessibile del metodo classico in quanto fa uso di due costanti, anche se sorge qualche problema in più nella scelta dei valori iniziali.

➤ *Metodo stagionale additivo*

Nelle vicinanze di  $n$  la serie storica è esprimibile come

$$L_n + T_n(t - n) + S_n \quad [25]$$

Dove  $S$  è un fattore stagionale di periodo  $s$ .

➤ *Metodo stagionale moltiplicativo*

Nelle vicinanze di  $n$  la serie storica è esprimibile come

$$L_n + T_n(t - n)S_n \quad [26]$$

### 6.6.6 Metodo di regressione

I metodi di regressione possiedono caratteristiche che ne consentono l'applicabilità in differenti situazioni e con riferimento a strutture di dati tra loro difformi. Per questo, la necessità di pervenire ad un modello soddisfacente implica necessariamente che vengano prese in considerazione più tipologie di metodi, e ciò con l'obiettivo di rendere il risultato della scelta il più in linea possibile con le esigenze legate al risultato richiesto.

Quando è possibile individuare delle relazioni tra le variabili di interesse, allora il problema della previsione si pone sotto un altro punto di vista rispetto a quello analizzato nei metodi precedenti. Infatti, un conto è tentare di prevedere l'andamento di una determinata serie storica nella sua dinamica temporale o spaziale, altro è tener conto dei legami esistenti tra variabili e stabilire il tipo di relazione che tra loro intercorre. Le previsioni, non necessariamente *time dependent*, saranno così espresse come funzione di una serie di fattori che influenzeranno il risultato finale. Tra i vantaggi di una tale impostazione vi è certamente quello di consentire una migliore comprensione del fenomeno e di lasciare spazio alla sperimentazione sui risultati conseguibili combinando in modo diverso i vari input ed osservando in che modo tali combinazioni agiscono sulla previsione (output).

Attraverso l'applicazione di modelli con variabili esplicative, inoltre, si apre la strada alla possibilità di influenzare il futuro sulla base di decisione prese oggi e di ottenere previsioni più accurate grazie al fatto che può essere stimata l'influenza degli input sugli output. A prescindere dal modello utilizzato (regressione semplice o multipla) è compito del previsore decidere quale o quali variabili costituiranno la previsione e quali svolgeranno invece da input del modello.

Nella regressione lineare, il modello assume che la variabile dipendente,  $y_i$  sia una combinazione lineare nei parametri ( $\beta$ ) senza però la necessità che sia lineare nella variabile indipendente.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{con } i=1, \dots, N \quad [27]$$

Nell'equazione  $x$  i termini  $\beta$  sono i parametri,  $x_i$  è la variabile indipendente ed  $\varepsilon_i$  è il termine di errore.

Il modello di regressione lineare semplice può quindi essere rappresentato nell'equazione

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad [28]$$



In questa equazione  $e_i$  rappresenta il residuo ed è dato dall'equazione

$$e_i = y_i - \hat{y}_i \quad [29]$$

Per la stima dei parametri  $\beta$  si utilizza quindi il metodo dei minimi quadrati che minimizzano l'errore dei residui.

La regressione multipla è caratterizzata da  $j$  variabili indipendenti e ciò comporta che la stima dei parametri con il metodo dei minimi quadrati è ottenuta tramite  $j$  equazioni normali. Da ciò si ha che la regressione lineare è rappresentata dalla seguente equazione:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_j x_{ji} + \varepsilon_i \quad [30]$$

Ed i residui sono dati dall'equazione:

$$e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \dots - \hat{\beta}_j x_{ji} \quad [31]$$

## **6.7 I modelli ARIMA**

Questa categoria di modelli lineari, fondamentali nell'ambito dell'analisi delle serie storiche, può essere descritta a partire dai processi stocastici stazionari secondo un approccio deduttivo. Da questo punto di vista è possibile partire dal teorema di Wold, la cui approssimazione è un modello MA (moving average) caratterizzato da un numero finito di autocovarianze solo se invertibili. Tuttavia, se un processo è invertibile allora è esprimibile in funzione delle variabili casuali passate ovvero nei termini di un processo AR (autoregressivo). Combinando insieme le due strutture lineari, si ottiene un modello misto noto come ARMA (autoregressive moving average). Il teorema di Wold afferma che ogni processo stocastico stazionario può decomporre nella somma di due processi stocastici tra loro incorrelati noti come componente deterministica e componente non-deterministica. La prima costituisce la parte puramente stocastica in quanto combinazione lineare di processi White Noise, l'altra è una combinazione lineare di onde periodiche, con ampiezza e fase che sono delle variabili casuali, ma con frequenza deterministica e pertanto costituiscono un'onda perfettamente periodica determinabile con esattezza. La determinazione esauriente della componente deterministica consente di focalizzare l'analisi statistica sulla sola componente stocastica.

L'identificabilità di un modello ARMA riguarda la possibilità, teorica, di partire dalla serie storica per arrivare, mediante un insieme di procedure statistiche, al processo stocastico che ha generato la serie stessa, il tutto all'interno di una classe "circonscritta" di metodi. Il primo passo consiste nell'ottenere

una stima della funzione di autocorrelazione globale e parziale mediante opportuni stimatori. Tali funzioni costituiscono l'elemento centrale dell'analisi in quanto, come è stato già accennato, se il riferimento è un modello ARMA stazionario e invertibile, allora è possibile determinare una corrispondenza biunivoca fra modello e funzione di autocorrelazione. Ma il modello ARMA e la funzione di autocorrelazione non esauriscono la conoscenza del processo stocastico, che invece richiede un'assunzione sulla distribuzione di probabilità delle variabili casuali che compongono il processo stesso. Si è quindi affermato come nel caso di processi stazionari e invertibili di tipo Gaussiano, la conoscenza della funzione di autocovarianza (e quindi della funzione di autocorrelazione) equivale alla conoscenza del processo. Queste considerazioni sono valide anche con riferimento a serie non stazionarie ma riconducibili al caso stazionario mediante l'utilizzo dell'operatore alle differenze finite. Se quindi si parte dalla reale possibilità di identificare un modello che sia una rappresentazione finita del processo stocastico, diventa interessante poter utilizzare una strategia che dai dati disponibili sia in grado di portare il ricercatore verso il miglior modello (costruzione del modello). Il riferimento è chiaramente alla nota procedura di Box e Jenkins che prevede essenzialmente lo svolgimento di quattro fasi prima di utilizzare il modello per i diversi fini (previsione, controllo, simulazione, etc.). Il primo step consiste nell'analisi preliminare che implica la verifica della normalità del White Noise, l'eventuale eliminazione degli outliers, la verifica della stazionarietà in media, varianza e autocovarianza. Nel primo caso è sufficiente verificare, attraverso test opportuni di accostamento alla distribuzione normale. Il secondo aspetto è legato alla necessità di eliminare l'eventuale asimmetria della distribuzione e/o tutti quei dati anomali che possono avere effetti rilevanti sulla varianza della serie e quindi sull'autocovarianza a tutti i lags. È questo un aspetto fondamentale per giungere ad una corretta identificazione di un modello ARMA. L'ultimo aspetto, che assume un ruolo centrale, riguarda per l'appunto il bisogno di costruire una serie stazionaria a partire da una serie che non lo è. In presenza di non stazionarietà in media, ad esempio, tra i metodi più diffusi per eliminare la componente di trend vi è quello della differenziazione. La non stazionarietà in varianza può essere affrontata mediante l'operatore logaritmo che, diminuendo il divario tra massimo e minimi, riduce la variabilità della serie.

Quando le condizioni di stazionarietà richieste per i modelli Box & Jenkins non sono presenti si possono avere due forme di non stazionarietà:

- esplosiva;
- omogenea.

Si ha la prima quando almeno una radice dell'equazione caratteristica è in modulo minore di 1. Si ha la seconda quando almeno una delle radici dell'equazione caratteristica è unitaria (cioè sul cerchio di raggio unitario).

I fenomeni socio-economici ben difficilmente presentano non stazionarietà esplosiva, limitandosi a forme omogenee, così dette perché a parte variazioni nel livello e/o nell'andamento di fondo (trend), la serie è di tipo stazionario. In altri termini la serie non è temporaneamente costante nel suo livello medio, ma comunque tende a disporsi stabilmente intorno a tale livello medio.

Un possibile modo di rappresentare una serie storica non stazionaria omogenea è introdurre in un modello ARMA un operatore alle differenze finite di ordine opportuno.

Integrando le componenti di Autoregressione (AR) e delle Medie Mobili (MA) con la componente alle differenze finite si ha il modello ARIMA.

Per definire formalmente tale modello si deve prima definire l'operatore autoregressivo generalizzato

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_{p+d} B^{p+d} \quad [32]$$

che è un polinomiale di grado  $p+d$  con  $d$  radici uguali ad 1 e le altre  $p$  radici maggiori di 1.

Pertanto:

$$\begin{aligned} \phi(B) &= (1 - r_1 B)(1 - r_2 B^2) \dots (1 - r_{p+d} B^{p+d}) \\ &= (1 - r_1 B)(1 - r_2 B^2) \dots (1 - r_p B^p)(1 - B)^d \end{aligned} \quad [33]$$

Questo perché  $d$  radici sono unitarie.

I fattori della parte destra dell'equazione meno l'ultimo sono niente altro che l'operatore  $\Phi(B)$  di un AR stazionario.

Quindi:

$$\phi(B) = \Phi(B)(1 - B)^d \quad [34]$$

Cioè:

$$\phi(B)z_t = \Phi(B)(1 - B)^d z_t = \Phi(B)\nabla^d z_t = \Phi(B)a_t \quad [35]$$

che scritto per esteso diviene:

$$z_t = \phi z_{t-1} - \dots - \phi_{p+d} z_{t-p-d} = a_t - \theta a_{t-1} - \dots - \theta_q a_{t-q} \quad [36]$$

$$\text{Pertanto se } w_t = \nabla^d z_t \quad [37]$$

L'equazione [35] diviene:

$$\Phi(B)w_t = \theta(B)a_t \quad [38]$$

che altro non è se non un ARMA sulle differenze di ordine  $d$  dei valori  $z_t$ .

Quindi sostituendo  $\nabla^d z_t$  con  $w_t$  il processo ARIMA sulla variabile  $z_t$  si riduce ad un ARMA sulla variabile  $w_t$ .

Allora il processo non stazionario  $\phi(B)$  è esprimibile come combinazione del processo stazionario  $\Phi(B)$  e dell'operatore alle differenze  $\nabla^d = (1-B)^d$ .

Tale combinazione determina il processo integrato ARIMA che pertanto è parte di una classe di processi più ampia di quelli ARMA che da essa discendono.

Si aggiunga infatti nel modello che esprime  $z_t$  una costante arbitraria a tutti i termini fino a quello di ordine  $t-1$ ; in altri termini:

$$z_t = (z_{t-1} + c) + \Phi_1 \{(z_{t-1} + c) - (z_{t-2} + c)\} + \dots + \Phi_p \{(z_{t-p} + c) - (z_{t-p-1} + c)\} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad [39]$$

che è come dire:

$$z_t = \left\{ z_{t-1} + \Phi_1 (z_{t-1} - z_{t-2}) + \dots + \Phi_p (z_{t-p} - z_{t-p-1}) + \right\} + c + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad [40]$$

Cioè con l'aumento di tutti i termini della costante  $c$ , anche  $z_t$  risulta aumentato di  $c$ . Quindi una serie non stazionaria ma omogenea si comporta come una serie stazionaria, poiché il suo andamento è indipendente dal livello della serie.

I modelli ARIMA, come è stato accennato, possono essere utilizzati per diversi fini ma è fuori discussione che il campo di applicazione che ha suscitato maggior interesse è stato proprio quello relativo alla previsione. In generale, la possibilità offerta dai modelli ARIMA di ricostruire il meccanismo probabilistico sottostante la serie osservata, fa sì che il suo uso predittivo risulti più potente se confrontato con il metodo regressivo che inoltre prevede una specificazione di tipo extra-statistico. Fatto interessante per quanto riguarda i modelli ARIMA è invece relativo alla circostanza che l'esplicitazione avviene proprio all'interno della metodologia statistica utilizzata per realizzare il modello e si basa direttamente sui dati osservati. Altro problema è quello relativo alla quantità di informazioni disponibili e all'esattezza dei parametri del modello, ma è possibile comunque affermare che nel caso di serie abbastanza lunghe l'errore che si può commettere è comunque trascurabile.

## 6.8 Group method of data handling - GMDH

Il Group method of data handling (GMDH) è una famiglia di algoritmi induttivi per la modellazione matematica di set di dati multi-parametri con ottimizzazione strutturale e parametrica del modello completamente automatica [30].

Il metodo è stato quindi utilizzato in campi come il data mining, forecasting, previsione, sistemi complessi di modellazione, ottimizzazione e pattern recognition.

Gli algoritmi GMDH attuano una procedura induttiva per la scelta di modelli polinomiali gradualmente complicati che forniscono la migliore soluzione per mezzo del cosiddetto criterio esterno. Un modello GMDH con più ingressi e un'uscita è un sottoinsieme di componenti della funzione base:

$$Y(x_1, \dots, x_n) = a_0 + \sum_{i=1}^m a_i f_i \quad [41]$$

dove  $f$  sono funzioni elementari dipendenti da diversi gruppi di ingressi,  $a$  sono coefficienti e  $m$  è il numero dei componenti funzionali di base.

Per trovare i migliori algoritmi il modello GMDH considera vari sottogruppi che compongono la funzione base chiamati modelli parziali. I coefficienti di questi modelli sono stimati con il metodo dei minimi quadrati. Il modello procede per auto-organizzazione, cioè aumentando gradualmente il numero di componenti del modello parziali fino a trovare una configurazione con complessità ottimale sulla base di un indicatore di valore minimo del criterio esterno.

L'algoritmo maggiormente utilizzato, che assume gradi di complessità sempre maggiori, è il polinomio di Kolmogorov-Gabor.

$$Y(x_1, \dots, x_n) = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_{ij} + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ijk} x_i x_j x_k + \dots \quad [42]$$

L'ampio sviluppo della teoria GMDH ha portato ad un ampio spettro di algoritmi ai quali corrispondono alcune condizioni specifiche ed un particolare tipo di applicazione.

Gli algoritmi GMDH possono quindi differire per il tipo di funzione elementare, il metodo di configurazione della struttura del modello e della sua complessità, i criteri esterni di verifica o il tipo di attività di modellazione. La scelta dell'algoritmo dipende dal livello di rumore dei dati, la loro completezza e se di tipo continuo o discreto. I primi algoritmi GMDH avevano adottato tre diversi tipi di funzioni elementari come, grafici probabilistici, formule di Bayes o teorema della probabilità delle cause e polinomi di secondo ordine. Tuttavia, il rapido sviluppo della teoria GMDH e l'ampio spettro degli algoritmi utilizzati hanno generato un approccio che vede due categorie di classificazione:

- GMDH con algoritmo parametrico;
- GMDH con algoritmo non parametrico.

Gli algoritmi parametrici sono prevalentemente utilizzati per descrivere i sistemi esatti, caratterizzati da dati rumorosi e varianza basse. Nel caso di sistemi mal definiti e con alta varianza e dati rumorosi, è invece giustificata l'applicazione di algoritmi non parametrici.

Gli algoritmi parametrici sono a sua volta stati classificati in base al tipo di funzione di attivazione o al tipo di complessità della struttura del modello. Per quest'ultima classificazione, si distinguono:

- *Algoritmi combinatori*, noti anche come algoritmi di auto-organizzazione mono-strato, che esegue una ricerca di modellizzazione tra tutti i metodi di studio;
- *Multistrato o algoritmi iterativi*, applicano una procedura iterativa che aumenta la complessità del modello mentre un criterio esterno identifica i metodi da implementare nello strato successivo. Questo tipo di approccio non esegue una verifica completa di tutti in metodi di studio, ciò garantisce un tempo di calcolo ridotto anche se il numero di variabili indipendenti oggetto del calcolo è maggiore.

Per quanto riguarda il tipo di funzione di attivazione, gli algoritmi GMDH si distinguono in polinomiali, armonici, moltiplicativi-additivi e fuzzy.

Tutti questi algoritmi saranno ampiamente discussa nei paragrafi successivi. La natura adattiva degli algoritmi GMDH ha permesso di operare alcune modifiche nella funzione GMDH di base. Grazie a questo è stato possibile applicare migliorie che rendano il modello applicabile con maggiore accuratezza a realtà peculiari.

### **6.8.1 Algoritmo combinatorio (COMBI)**

L'algoritmo GMDH combinatorio classico genera modelli a partire dalla combinazione di tutte le possibili variabili di input e seleziona il migliore modello finale dal set generato dei modelli secondo il criterio di selezione esterno scelto. Ivakhnenko et al. [31] hanno indicato algoritmo combinatorio come un metodo completo di induzione matematica. L'algoritmo combinatorio ordina i modelli gradualmente, aumentando i termini da 1 ad  $n$  (cioè il numero di argomenti) mentre un criterio esterno indicherà la soluzione ottimale tra il set di modelli con stessa complessità. Pur tuttavia, il criterio esterno ha un peso minimo nel piano della complessità rispetto al criterio di selezione del tipo di algoritmo. Lo svantaggio principale degli algoritmi combinatori è essenzialmente correlato alla loro richiesta di capacità computazionale del computer. Infatti, grazie alla complessa procedura il tempo di calcolo è elevato, pertanto si dovrà scegliere un limitato numero di termini in input affinché i tempi di elaborazione risultino contenuti. Sempre Ivakhnenko et al. [31] propongono diversi modi per aumentare il numero di termini nel modello finale. Uno di questi è noto come il metodo di

troncamento del triangolo di smistamento secondo il quale ogni modello con più di un numero specificato di termini viene scartato dalla procedura di smistamento. Ovviamente, questo metodo viola la proprietà di ricerca esaustiva del modello ottimale. Successivamente l'aumento del dominio di ingresso è stato superato selezionando a caso gli argomenti di input come coppie. Un inconveniente ulteriore legato all'applicazione di algoritmi combinatori, ma comune a tutti gli algoritmi GMDH polinomisli, è dato dalle stime distorte dei coefficienti a causa del criterio esterno basato sul metodo dei minimi quadrati. Ivakhnenko et al. [32] sostengono che il metodo delle variabili strumentali come criterio esterno potrebbe sostituire quello dei minimi quadrati producendo minori errori di stima.

I vantaggi del monostrato auto-organizzante potrebbero essere riassunte nella loro semplicità e la loro capacità di svolgere una completa analisi dei set di modelli elaborati. Tuttavia, se si aumenta il numero di termini, e quindi la complessità del modello, deve anche aumentare il numero di osservazioni a disposizione. Ciò alla fine limita l'applicazione dell'algoritmo a data set con input completi e numerose osservazioni disponibili.

### 6.8.2 Algoritmi multistrato

L'algoritmo GMDH multistrato ha una struttura molto simile a quella di *multilayer feedforward neural networks* ma il numero di strati così come il numero di nodi è oggettivamente definito da un criterio esterno. Osservando la Figura 6.4 è chiaro che nel primo strato il numero di nodi è pari al numero di ingressi mentre per gli strati successivi è uguale al numero di coppie di variabili per il vettore caratteristica (cioè variabili di input). Il numero di layer nascosti può essere specificato come input o determinato secondo un valore di soglia del criterio esterno.

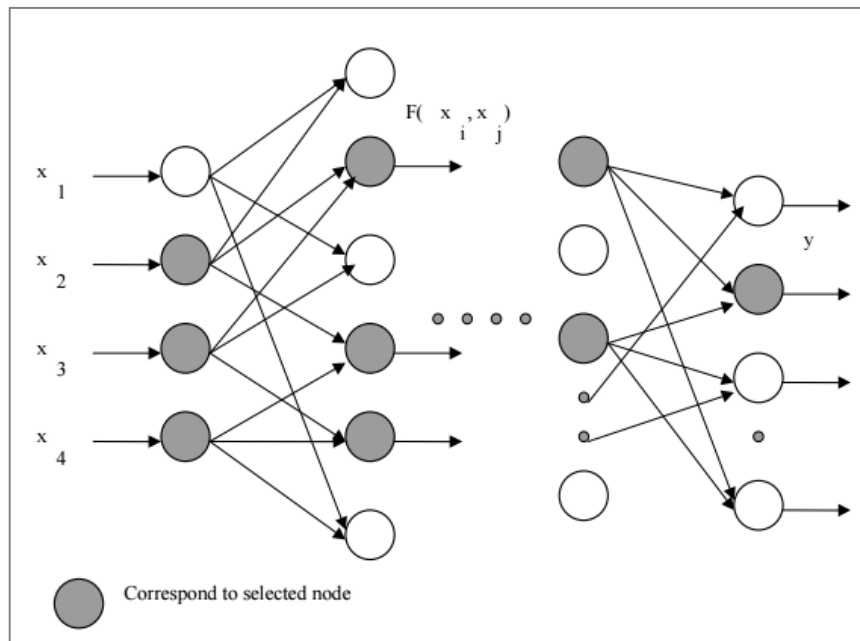


Figura 6.4 - GMDH Algoritmo multistrato.

Gli algoritmi multistrato non svolgono una ricerca esaustiva tra tutti i set di modelli, ma se il numero di modelli selezionati in ogni layer è abbastanza grande, verrà sempre scelta la soluzione ottimale. Il tipo più popolare di funzione di attivazione è il polinomio di secondo grado, tuttavia l'applicazione di diversi tipi di funzioni di attivazione ha permesso l'applicazione del modello GMDH in diversi campi di ricerca.

Per quanto rappresentato, gli algoritmi multistrato risultano superiori agli algoritmi combinatori per via che possono elaborare un numero maggiore di termini nel modello finale. Questo tuttavia comporta che negli algoritmi multistrato vi è una crescita esponenziale del volume di computazione all'aumentare del numero di variabili del data set. La loro applicazione risulta quindi suggerita per i casi in cui il data set ha input ridotti e mal definiti. L'algoritmo genera quindi una singola equazione di uscita che prevede la conoscenza dell'output. Tale inconveniente è stato superato progettando un algoritmo multistrato per sistemi con multi-uscita. L'unica condizione è che le variabili di output dovrebbero essere note a priori. A tal proposito si sostiene che è meglio sotto-determinare il sistema piuttosto che sopra-determinarlo, in questo modo si garantisce la determinazione delle leggi che governano il processo oggetto dello studio anche se con precisione inferiore.

Una caratteristica distintiva di questo algoritmo è l'introduzione di un ulteriore criterio a ogni livello (cioè l'equilibrio delle variabili di criterio), che prende in considerazione il set di previsione e decide sul set di equazioni. Pertanto, indagando il set teorico di algoritmi multistrato, ciò che si deve verificare nell'indagine sono essenzialmente due proprietà. In primo luogo, la ricerca nel set di strutture deve procedere nella direzione di aumentare la complessità e il massimo grado di termini e in secondo luogo per qualsiasi struttura ammissibili deve essere il fattore che conduce ad esso dagli strati precedenti. L'algoritmo "Precisione", che è un algoritmo multistrato a due stadi, soddisfa queste proprietà e attua nella prima fase una ricerca limitata a monomi con crescenti esponenti. Nella seconda, invece, la ricerca viene estesa a tutti i set di modelli aventi uno o due termini all'interno dell'insieme di descrizioni parziali.

È quindi possibile classificare gli algoritmi multistrato in quelli con calcolo dei residui dopo ogni strato di selezione e quelli senza calcolo dei residui.

Gli algoritmi multistrato sono stati negli anni applicati in un ampio spettro di applicazioni e diverse modifiche sono state progettate per aumentare la loro accuratezza.

### **6.8.3 Algoritmo Fuzzy**

Nel metodo convenzionale GMDH lo scostamento tra i valori osservati dell'output e la sua stime si suppone abbia una distribuzione gaussiana, pertanto viene utilizzata un tipo di analisi basata sulla



regressione lineare per stimare i parametri. Tuttavia, tale ipotesi è spesso violata e il metodo dei minimi quadrati non risulta appropriato all'analisi. Inoltre, la maggior parte dei sistemi reali sembra seguire il principio di incompatibilità introdotto da Zadeh e la *teoria Fuzzy* risulta un approccio più adeguato per il procedimento di modellazione.

Si è quindi tentato di rappresentare le relazioni tra input-output con modelli in cui i parametri sono riconducibili a numeri fuzzy che fanno parte di un insieme  $A$  costituito da coppie ordinate di un generico elemento  $u$  e del suo grado di appartenenza:

$$A = \{(u, \mu_A(u)) : u \in U\} \quad [43]$$

La struttura dell'algoritmo rimane la stessa, ma i parametri Fuzzy sono utilizzati nelle descrizioni parziali e vengono ricercati attraverso il metodo di regressione lineare. Questo tipo di approccio ha fatto sì che venisse elaborato l'algoritmo Fuzzy GMDH che ha una struttura simile a quella degli algoritmi convenzionali ma con l'applicazione delle regole Fuzzy.

#### **6.8.4 Algoritmo non parametrico**

Lo sviluppo di algoritmi non parametrica è stato motivato dalla debolezza riscontrata negli algoritmi parametrici. L'approccio con questa tipologia di algoritmi permette quindi di superare le difficoltà causate da una grande dimensione di dati iniziali e l'utilizzo di criteri interni al modello e di clustering, aumentando di fatto l'affidabilità dei coefficienti senza la divisione dei campioni di dati in più gruppi. Tale tipologia di algoritmi è utilizzata quindi per le analisi di clusterizzazione, che è un metodo molto soddisfacente per le previsioni a lungo periodo.

La clusterizzazione è un metodo matematico di modellizzazione debole, che in accordo con la legge di adeguatezza di un oggetto, è la tecnica più appropriata per la modellazione di sistemi complessi del mondo reale. L'analisi tipo cluster è un problema tridimensionale in cui deve essere ricercata la larghezza del cluster, la lunghezza e i loro numeri. Generalmente, questo problema è risolto tramite un procedimento di ordinamento deduttivo o induttivo. I metodi deduttivi sono soggettivi, in quanto sono completamente basati sulle istruzioni da parte della persona che effettua il raggruppamento. I metodi induttivi, invece, poiché si basano sul procedimento di auto-organizzazione, sono oggettivi e quindi più adeguati alla modellizzazione con le caratteristiche del data set sopra specificato.

Ritornando all'algoritmo non parametrico, dato un campione di osservazioni

$$\{(x_i, y_i)\}_{i=1}^n \quad [44]$$

la sottostante relazione fra variabili è del tipo  $y_i = m(x_i) + \varepsilon_i$

L'obiettivo dell'analista è quello di stimare la funzione  $m$  a partire dal campione osservato, in modo da poter conoscere la relazione che lega  $x$  ed  $y$ , sia per scopi descrittivi che per scopi predittivi. Se per un punto fissato  $x$  ci sono osservazioni ripetute, allora la stima di  $m(x)$  può essere fatta usando semplicemente la media dei valori  $y$  corrispondenti. Purtroppo, nella maggior parte dei casi pratici tale situazione non si presenta, per cui la stima deve essere fatta seguendo altre procedure.

In generale, vengono considerate procedure di media locale, ossia si stima  $m(x)$  attraverso una media aritmetica ponderata delle osservazioni  $y$  corrispondenti ai valori  $i$  appartenenti ad un certo intorno del punto  $x$  di riferimento. Per cui, considerando il punto  $x$ , il valore che la funzione  $m$  assume in corrispondenza di tale punto viene stimata nel seguente modo:

$$\hat{m}(x) = n^{-1} \sum_{i=1}^n W_{ni}(x) y_i \quad [45]$$

dove l'insieme  $\{(W_{ni}(x))\}_{i=1}^n$  denota una sequenza di pesi che può dipendere dall'intero vettore  $\{X_i\}_{i=1}^n$ .

Lo stimatore  $m(x)$  è usualmente denominato *smoother*, mentre la procedura di stima è chiamata *smooth*. Nello stimatore  $m(x)$  un ruolo chiave è giocato dal sistema di pesi  $\{(W_{ni}(x))\}_{i=1}^n$ ; i valori di tale sequenza regolano l'ampiezza dell'intorno di  $x$  impiegato per calcolare la media ponderata locale delle  $y$ . Una media locale su un intorno troppo largo determina uno stimatore  $m$  con bassa varianza, ma distorto; dall'altra parte, una media locale calcolata su un intorno troppo piccolo implica la presenza di una stima corretta, ma con alta variabilità. Pertanto, si rende necessario un compromesso che bilanci il trade-off fra i due casi.

Se si suppone che i pesi  $\{(W_{ni}(x))\}_{i=1}^n$  siano positivi e che, per ogni  $x$ , la loro somma soddisfi la relazione:

$$n^{-1} \sum_{i=1}^n W_{ni}(x) = 1 \quad [46]$$

allora  $m(x)$  è la soluzione del metodo di stima dei minimi quadrati al punto  $x$ , se è la soluzione del seguente problema di minimizzazione:

$$\min_{\theta} n^{-1} \sum_{i=1}^n W_{ni}(x) (y_i - \theta)^2 = \min \left( n^{-1} \sum_{i=1}^n W_{ni}(x) (y_i - \hat{m}(x))^2 \right) \quad [47]$$

### 6.8.5 Algoritmi semiparametrici

Molto spesso, per sfruttare i vantaggi sia dell'approccio parametrico che di quello non parametrico, si considerano modelli che sono parzialmente specificati e che sono detti semiparametrici.

Un modello semiparametrico è del tipo:

$$y = f(z, \theta) + m(x) + \varepsilon \quad [48]$$

dove sia  $z$  che  $x$  sono variabili esplicative e  $\theta$  è l'insieme dei parametri da cui dipende la funzione  $f$ , che viene completamente specificata. Nell'equazione precedente, si nota che la parte deterministica del modello è formata da due componenti:

- $f(z; \theta)$  è la parte parametrica del modello, in quanto la funzione  $f$  è specificata a priori e dipende da un certo insieme di parametri  $\theta$ ;
- $m(x)$  è la parte non parametrica del modello, che verrà stimata con una delle tecniche non parametriche descritte in precedenza.

## 6.9 Misure di bontà di adattamento e di accuratezza della previsione

In molte analisi statistiche in cui un modello viene adattato a un set di dati, il termine accuratezza si riferisce alla capacità del modello di riprodurre i dati sui quali è stato stimato e cioè della bontà di adattamento del modello. L'accuratezza nella previsione, invece, misura la capacità del modello a riprodurre i dati futuri della serie.

Per misurare la bontà di adattamento/previsione sono più comunemente utilizzati i seguenti indici:

- *errore medio* (mean error: ME): media aritmetica degli errori;

$$ME = \frac{1}{m} \sum_{t=1}^m f_t \quad [49]$$

- *errore quadratico medio* (mean squared error: MSE): media aritmetica dei quadrati degli errori;

$$MSE = \frac{1}{m} \sum_{t=1}^m f_t^2 \quad [50]$$

- *errore medio assoluto* (mean absolute error: MAE): media aritmetica degli errori presi in valore assoluto;

$$MAE = \frac{1}{m} \sum_{t=1}^m |f_t| \quad [51]$$

- *errore medio assoluto percentuale* (mean absolute percentage error: MAPE): media aritmetica degli errori relativi, presi in valore assoluto e moltiplicati per 100.

$$MAPE = \frac{1}{m} \sum_{t=1}^m \frac{|f_t|}{y_t} \cdot 100 \quad [52]$$

ME è l'unico indice che può assumere anche valori negativi. Il MAE avrà valori bassi perché elementi positivi e negativi, nella sommatoria, tenderanno a compensarsi. Il segno di ME fornisce informazioni se si tende, in media, a sopravvalutare ( $ME < 0$ ) o sottovalutare ( $ME > 0$ ) il fenomeno. MSE e MAE rendono positivi i singoli addendi della sommatoria. Per la presenza dell'esponente, MSE è meno agevole da interpretare da persone non specialiste.

I tre indici ME, MSE e MAE forniscono un valore che dipende dall'unità di misura della serie. Il MAPE elimina questo problema in quanto l'errore viene relativizzato dividendolo per il valore osservato. Il MAPE può però essere usato solo se il fenomeno è misurabile su scala a rapporto. Per il calcolo del MAPE sorgono difficoltà quando la serie osservata contiene valori nulli o molto prossimi a zero.

Uno dei problemi interpretativi dei valori degli indici sopra esposti concerne il fatto che questi indici esprimono funzioni di distanza, fra valori osservati e previsti, che possono essere a volte molto diverse fra loro ed anche discordanti. Occorre quindi decidere in via preliminare l'indice da usare per misurare l'accuratezza della previsione. Al fine di valutare le prestazioni del metodo di previsione, è buona regola condurre anche analisi grafiche degli errori, per esaminare il loro andamento rispetto al tempo.

## CAPITOLO 7 *Le reti neurali*

### 7.1 Introduzione alle reti neurali

Le reti neurali artificiali (Neural Networks) sono nate per riprodurre attività tipiche del cervello umano come la percezione di immagini, il riconoscimento di forme, la comprensione del linguaggio, il coordinamento senso-motorio, etc.. Esse prendono spunto dalle caratteristiche del cervello umano e dall'interazione che avviene a livello del singolo neurone e in gruppi di essi. Nel sistema nervoso esistono miliardi di *neuroni* (cellule nervose).

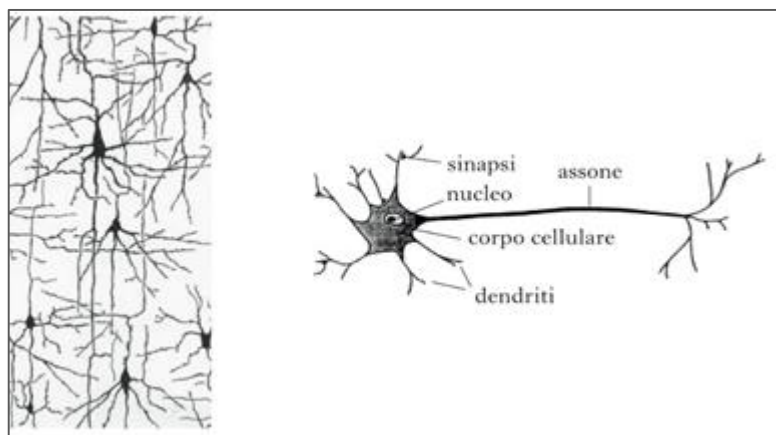


Figura 7.1 - Rappresentazione del neurone e dello strato di neuroni.

Un neurone biologico è composto da un corpo cellulare o “soma” dal quale partono molti collegamenti, chiamati “dendriti”, che ricevono segnali da altri neuroni, e da un collegamento di uscita, chiamato “assone”, con il quale il neurone trasmette informazioni ad altri neuroni. Ogni neurone ha una soglia di attivazione caratteristica: se i segnali provenienti da altri neuroni la superano, il neurone si attiva e trasmette un segnale elettrico sull’assone che arriva agli altri neuroni. Fra assone e dendrite esiste una sottile intercapedine, detta “sinapsi”, che permette la trasmissione del segnale attraverso un processo elettrochimico. Lo spessore della sinapsi nel tempo può variare, rafforzando o indebolendo il collegamento tra due neuroni. È dall’insieme dei valori di attivazione di tutti i neuroni che viene rappresentato il contenuto informativo momentaneo del cervello; la memoria invece è rappresentata dai valori di collegamento (più o meno forte) di tutte le sinapsi.

Le caratteristiche fondamentali del cervello sono la sua plasmabilità che deriva dal fatto che le sinapsi possono modificarsi nel tempo interagendo con segnali dal mondo esterno e la scomposizione

dell'informazione in più informazioni elementari contenute in ogni singolo neurone. Non è ancora chiaro il meccanismo di apprendimento del cervello umano ma è chiaro che il rafforzamento e l'indebolimento dei collegamenti sinaptici determinano la memorizzazione delle informazioni. Come per il cervello, in una rete neurale le informazioni sono scomposte in informazioni “elementari” contenute all'interno di ogni singolo neurone. Una rete neurale può essere vista come un sistema in grado di dare una risposta ad una domanda o fornire un output in risposta ad un input.

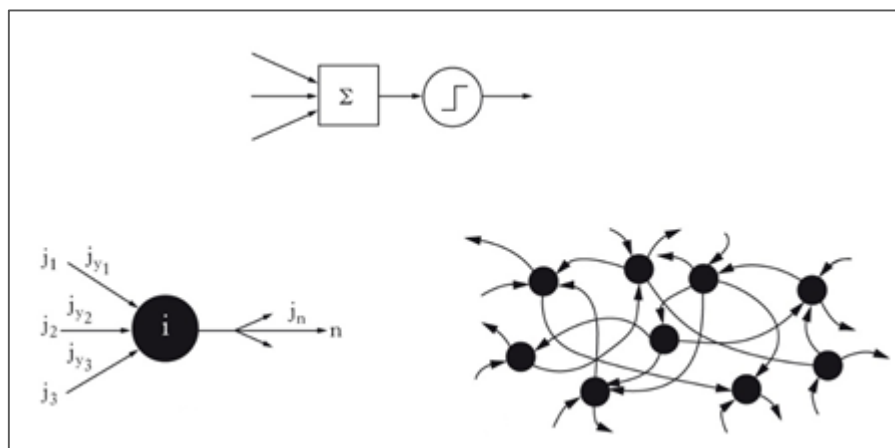


Figura 7.2 - Rappresentazione del neurone e dello strato di neuroni rappresentativi delle reti neurali artificiali.

La combinazione in/out, ovvero la funzione di trasferimento della rete, non viene programmata ma viene ottenuta attraverso un processo di “addestramento” con dati empirici. La rete apprende la funzione che lega l’output con l’input attraverso le presentazioni di esempi corretti di coppie (input, output). Per ogni input presentato alla rete, durante il processo di apprendimento, la rete fornisce un output che si discosta di una certa quantità dall’output desiderato: l’algoritmo di addestramento allora modifica alcuni parametri della rete e li avvicina ai valori ottimali per la soluzione dell’esempio; in questo modo l’algoritmo cerca di “accontentare” tutti gli esempi un po’ per volta. I parametri di cui si parla sono essenzialmente i pesi o fattori di collegamento tra i neuroni che compongono la rete. Una rete neurale è infatti composta da un certo numero di neuroni collegati tra loro da collegamenti “pesati”, proprio come lo sono i neuroni del cervello umano. Le reti neurali hanno caratteristiche sorprendentemente simili a quelle del cervello umano, in termini di capacità di apprendimento, scarsa precisione associata ad alta elasticità di interpretazione dell’input e quindi scarsa capacità di estrapolazione. In modo particolare l’alta elasticità di interpretazione dell’input, comunemente chiamata “resistenza al rumore” o “capacità di comprendere i dati rumorosi” è ciò che rende le reti neurali preferibili rispetto ad un comune sistema di elaborazione. Infatti, un sistema programmato ha bisogno di un input ben preciso per dare una risposta corretta, mentre una rete neurale è in grado di

dare una risposta abbastanza corretta ad un input parziale o impreciso rispetto a quelli utilizzati negli esempi di aggiornamento.

## 7.2 Il neurone artificiale

La schematizzazione di Figura 7.3 del neurone artificiale mette in evidenza come esso sia caratterizzato da un insieme di sinapsi, che corrispondono ai terminali di altri neuroni, da una soglia ( $b_i$ ) e da una funzione di attivazione ( $\Phi(A_i)$ ) [33].

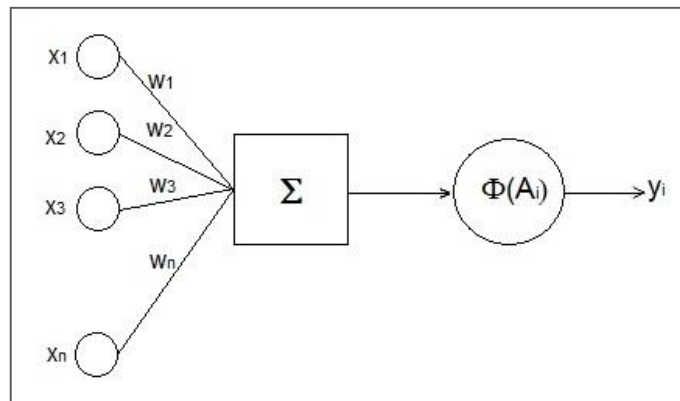


Figura 7.3 - Neurone artificiale.

Ciascun neurone elabora una somma pesata degli input che riceve e si attiva quando la quantità totale del segnale ricevuto supera la propria soglia di attivazione emettendo così un segnale che raggiunge tutte le unità connesse. Il potenziale di attivazione  $A_i$  di un neurone è dato da un input netto, ovvero la somma algebrica dei prodotti di tutti i segnali di ingresso e i valori dei pesi delle sinapsi, a cui viene sottratto il valore di soglia  $b_i$  del neurone, come espresso dalla seguente formula:

$$A_i = \sum_{j=1}^N w_{ij} x_j - b_i \quad [53]$$

La risposta del neurone  $y_i$  è calcolata sottoponendo il potenziale di attivazione così ottenuto all'azione di una funzione di attivazione  $\Phi(A)$ :

$$y_i = \Phi(A_i) = \Phi(\sum_{j=1}^N w_{ij} x_j - b_i) \quad [54]$$

Dove  $N$  indica il numero di unità dello strato a monte connesse al neurone  $i$ ;  $x_j$  rappresenta l'input  $i$ -esimo proveniente da altre unità o dall'ambiente esterno;  $w_{ij}$  rappresenta il peso della connessione tra il neurone  $i$  ed il neurone  $j$  e può assumere valori positivi o negativi;  $b_i$  è il valore di soglia, detto anche unità di bias, ed è equivalente ad un peso sinaptico di valore  $b$ , collegato ad un'ulteriore unità di input e che quindi può essere modificato con le stesse regole di apprendimento impiegate per gli altri pesi

sinaptici della rete. Sotto l'ipotesi di considerare, dunque,  $b_i$  come un'ulteriore unità di input, con attivazione costante e pari a -1, possiamo esprimere l'attivazione del neurone semplicemente come:

$$y_i = \Phi(\sum_{j=0}^N w_{ij} x_j) \quad [55]$$

facendo partire la sommatoria da zero e fissando  $x_0$  pari a -1.

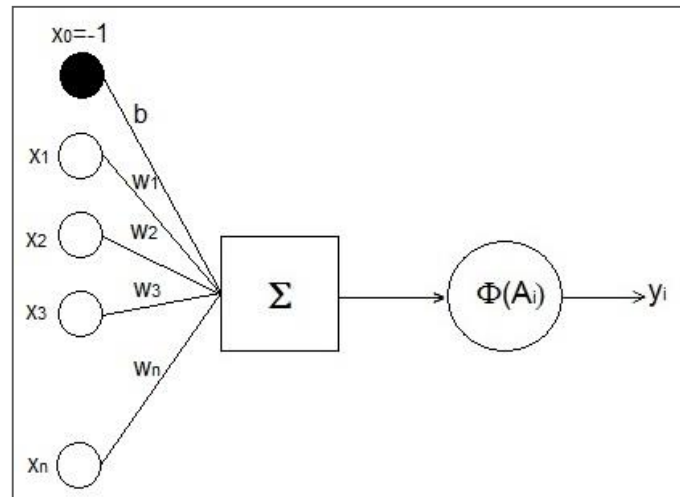


Figura 7.4 - Neurone Artificiale più Unità di Bias (cerchio nero).

### **7.3 Funzioni di attivazione**

Per quanto riguarda la funzione di attivazione esistono diversi tipi di funzioni. Nel 1943 McCulloch e Pitts proposero il primo modello matematico di rete neurale, riferito ad un neurone, rappresentato da una funzione binaria, con ingresso e uscita binari; in tale modello la soglia viene mantenuta fuori dal calcolo del potenziale di attivazione e la risposta del neurone è data da una funzione <<a gradino>>.

$$\Phi(A) = \begin{cases} 1 & \text{se } \sum_{j=1}^N w_{ij} x_i > b \\ 0 & \text{altrimenti} \end{cases} \quad [56]$$

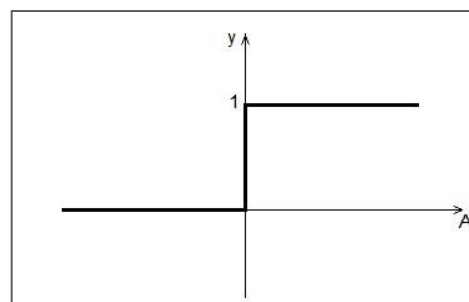


Figura 7.5 - Funzione di attivazione a gradino.



Funzione bipolare:

$$\Phi(A) = \begin{cases} 1 & \text{se } \sum_{j=1}^N w_{ij} x_i > b \\ -1 & \text{altrimenti} \end{cases} \quad [57]$$

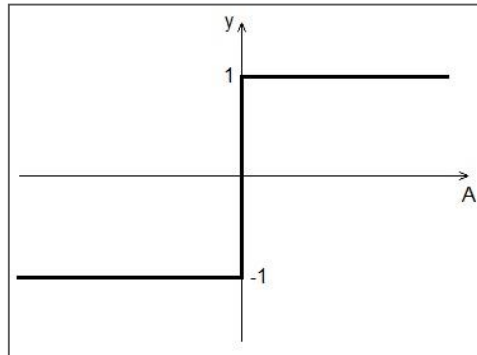


Figura 7.6 - Funzione bipolare.

Anche in questo caso il neurone può assumere solo due stati, risultando attivo o non attivo a seconda del livello di soglia.

Funzione lineare continua:

$$\Phi(A) = kA \quad [58]$$

dove  $k$  è una costante. In alcune situazioni questa funzione può essere forzata ad operare entro un certo intervallo (ad esempio nell'intervallo  $[0,1]$  o  $[-1,1]$ ) per contenere l'attivazione del neurone. Le funzioni continue permettono al neurone di trasmettere una gradazione di segnali di varia intensità che può essere opportunamente sfruttata dai neuroni precedenti.

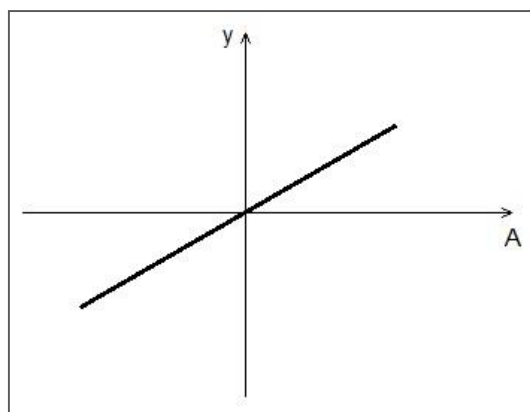


Figura 7.7 - Funzione lineare continua.

Funzione sigmoidea (fa parte delle funzioni continue non-lineari ed è una delle più utilizzate):

$$\Phi(A) = \frac{1}{1+e^{-kA}} \quad [59]$$

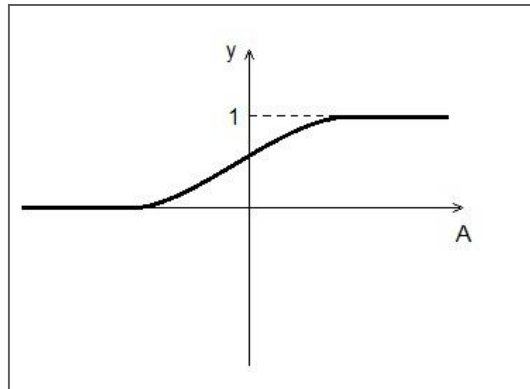


Figura 7.8 - Funzione sigmoidea.

$k$  rappresenta una costante che controlla l'inclinazione della curva; per  $k \rightarrow \infty$  la funzione sigmoide approssima quella a gradino. Le rette  $y = 0$  e  $y = 1$  sono asintoti orizzontali per la funzione sigmoide.

Funzione tangente iperbolica:

$$\Phi(A) = \frac{1-e^{-kA}}{1+e^{-kA}} \quad [60]$$

Molto simile alla funzione sigmoide, che però ha come asintoti orizzontali le rette  $y = 1$  e  $y = -1$ ;

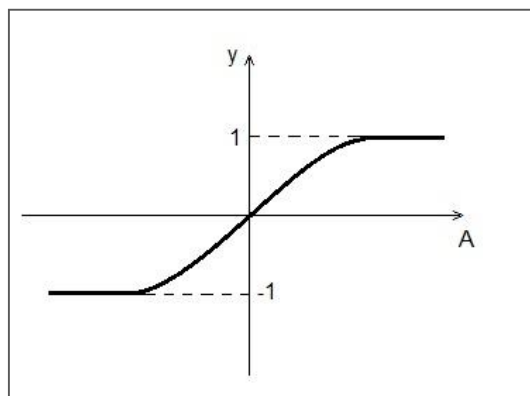


Figura 7.9 - Tangente iperbolica.

Nella maggior parte dei modelli tutti i neuroni della rete, eccetto i neuroni di ingresso (recettori), utilizzano la stessa funzione di attivazione per calcolare il proprio segnale di uscita.

## **7.4 Tipologie di reti neurali**

Esistono diversi tipi di reti neurali che si differenziano per alcune caratteristiche fondamentali:

- Tipo di utilizzo;
- Tipo di apprendimento;
- Algoritmo di apprendimento;
- Architettura dei collegamenti.

## **7.5 Modalità di utilizzo**

Dal punto di vista del tipo di utilizzo possiamo distinguere tre categorie fondamentali:

- *Memorie associative*: possono apprendere associazioni tra patterns (insieme complesso di dati come un insieme dei pixels di un'immagine) in modo che la presentazione di un pattern A dia come output il pattern B anche se il pattern A è impreciso o parziale (resistenza al rumore).
- *Simulatori di funzioni matematiche complesse (e non conosciute)*: sono in grado di comprendere la funzione che lega output con input in base a degli esempi forniti in fase di apprendimento. Dopo la fase di apprendimento, la rete è in grado di fornire un output in risposta ad un input anche diverso da quelli usati negli esempi di addestramento. Ne consegue una capacità della rete di interpolazione ed estrapolazione sui dati del training set. Tale capacità è facilmente verificabile addestrando una rete con una sequenza di dati input/output proveniente da una funzione nota e risulta, invece, utile proprio per il trattamento e la previsione di fenomeni di cui non sia chiaro matematicamente il legame tra input e output. In ogni caso la rete si comporta come una “*black box*”, poiché non svela in termini leggibili la funzione di trasferimento che è contenuta al suo interno. Di questo tipo fa parte la rete a retropropagazione dell'errore o *error back propagation* che è quella attualmente più utilizzata per efficacia e flessibilità.
- *Classificatori*: con essi è possibile classificare dei dati in specifiche categorie in base a caratteristiche di similitudine. In questo ultimo tipo di rete esiste il concetto di apprendimento non supervisionato o “auto-organizzante”, nel quale i dati di input vengono distribuiti su categorie non predefinite.

## 7.6 Modalità di apprendimento

Per quanto riguarda le modalità di apprendimento questo può distinguersi in *Apprendimento Supervisionato* e *Apprendimento non- supervisionato*. In generale l'apprendimento, come già accennato, consiste in una ricerca del set ottimale di pesi che realizza un buon adattamento della relazione tra input e output in modo che la rete impari a fornire risposte appropriate per ciascuno stimolo di ingresso.

Apprendimento supervisionato: la modifica dei valori sinaptici avviene impiegando una misura di errore tra la risposta fornita dalla rete neurale e la risposta desiderata per ogni vettore di input.

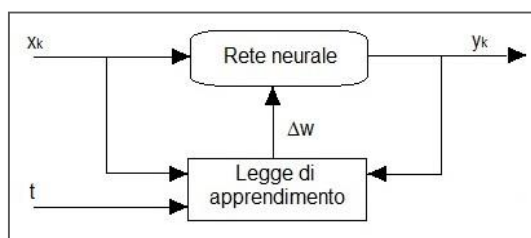


Figura 7.10 - Schema semplificato di apprendimento supervisionato.

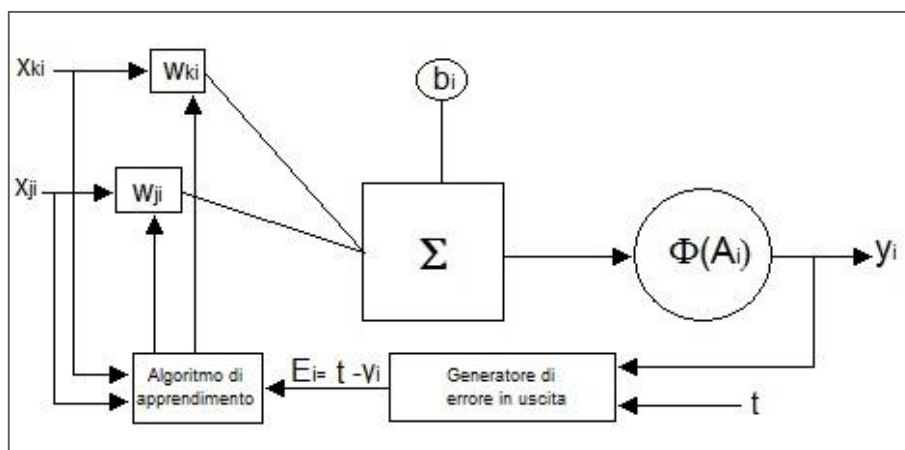


Figura 7.11 - Schema apprendimento supervisionato.

L'addestramento viene eseguito su un gruppo di coppie di pattern, ciascuna formata dal pattern di input  $x_k$  e dal pattern di risposta desiderata  $t$  ("target"). L'apprendimento include anche una gamma di algoritmi che richiedono solamente una misura di bontà della risposta della rete neurale, piuttosto che la specificazione della risposta esatta per ogni pattern di addestramento.

Apprendimento non supervisionato: questo tipo di apprendimento è basato su algoritmi d'addestramento che modificano i pesi della rete facendo esclusivamente riferimento ad un insieme di dati che include le sole variabili d'ingresso. La rete si auto-organizza durante l'esposizione dei pattern di input in modo da essere in grado di estrarre delle informazioni di similitudine tra di essi al fine di classificarli in categorie.

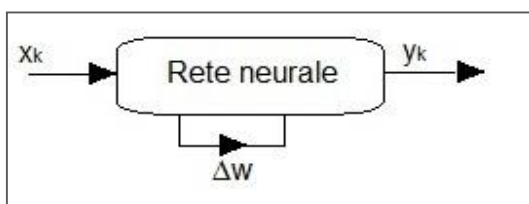


Figura 7.12 - Schema semplificato di apprendimento non supervisionato.

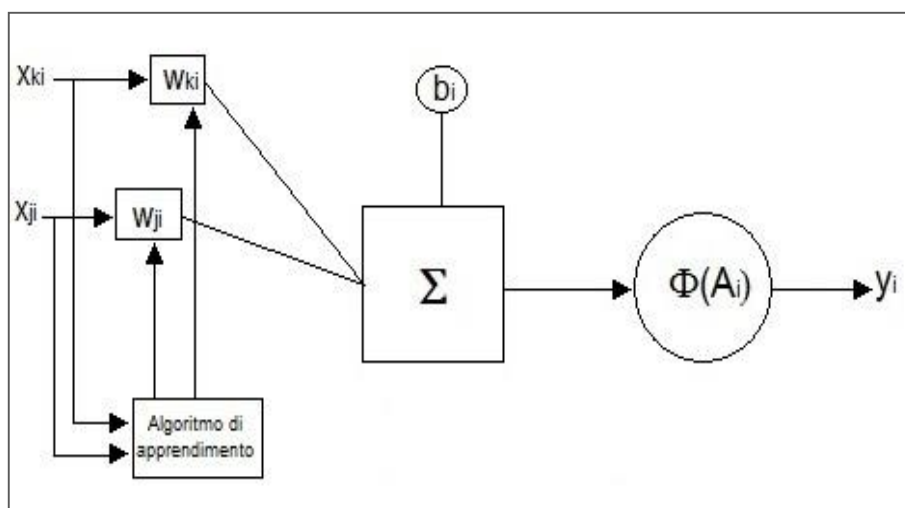


Figura 7.13 - Schema di apprendimento non supervisionato.

Tutti gli algoritmi di apprendimento presentano comunque delle caratteristiche generali comuni, come per esempio che i valori iniziali dei pesi sinaptici della rete vengono assegnati in modo casuale entro un piccolo campo di variazione (ad esempio  $[-0,1 ; 0,1]$ ) oppure vengono fissati tutti a zero. L'apprendimento consiste nella presentazione ripetuta di una serie di vettori, detti anche pattern di addestramento. Nell'apprendimento supervisionato ogni pattern è composto da una coppia, il vettore di ingresso e quello della risposta desiderata; nell'apprendimento auto-organizzato invece vi sono solo i vettori di ingresso; la modifica dei valori sinaptici  $\Delta w_{ik}$  della rete viene calcolata dopo ogni presentazione di un singolo pattern (<<apprendimento per cicli>>) oppure solo alla fine della presentazione di tutti i pattern di addestramento (<<apprendimento per epoche>>). La nuova

configurazione di valori sinaptici dopo un ciclo (o epoca) di addestramento è calcolata addizionando la modifica ottenuta  $\Delta w_{ik}^t$  alla configurazione sinaptica precedente  $w_{ik}^{t-1}$ :

$$w_{ik}^t = w_{ik}^{t-1} + \Delta w_{ik}^t \quad [61]$$

Gli algoritmi di apprendimento riguardano solo il calcolo di  $\Delta w_{ik}^t$ .

L'apprendimento riguarda dunque la sovrapposizione di nuove conoscenze  $\Delta w_{ik}^t$  su una base già consolidata di conoscenze precedenti  $w_{ik}^{t-1}$ . Per evitare il pericolo che le nuove conoscenze stravolgano o cancellino quanto già appreso, l'apprendimento procede in modo ricorsivo, cioè ciascun pattern viene presentato più volte alla rete neurale, e graduale ossia solo una frazione della modifica sinaptica viene effettivamente sommata ai valori sinaptici; dunque tale frazione detta “tasso di apprendimento”  $\eta$  è una costante che regola la velocità di apprendimento.

$$w_{ik}^t = w_{ik}^{t-1} + \eta \Delta w_{ik}^t \quad \text{con } 0 < \eta \leq 1 \quad [62]$$

Una volta che la fase di apprendimento è stata completata, i valori sinaptici vengono registrati ed è possibile studiare la risposta della rete su dei valori di test. La fase di test consiste nella presentazione di nuovi valori di pattern di ingresso e nel calcolo dell'attivazione dei nodi della rete senza però modificarne i pesi sinaptici. È possibile così studiare la capacità di generalizzazione a nuovi stimoli. Questa capacità di apprendere della rete è forse la sua caratteristica di maggiore attrazione perché permette di impiegare la rete neurale per risolvere problemi senza dover individuare direttamente la soluzione analitica, ma semplicemente esponendo il modello neurale ad una serie di esempi. Per questo motivo le reti neurali trovano vasta gamma di applicazione in problemi caratterizzati da trasformazioni non lineari, per cui è difficile (o avvolte impossibile) trovare soluzioni ottimali con i metodi analitici.

I metodi di valutazione variano a secondo del tipo di apprendimento utilizzato. Se si tratta di apprendimento supervisionato, si utilizza l'errore medio fra la risposta desiderata e la risposta fornita dalla rete neurale per ciascun pattern di addestramento: quando l'errore raggiunge un livello minimo prefissato i valori sinaptici vengono registrati e si passa alla fase di test. Il livello minimo di errore dipende da molti fattori come ad esempio il grado di precisione che si vuole raggiungere, la funzione di attivazione utilizzata ed il numero di unità di output. Per l'apprendimento non supervisionato il criterio cambia perché non esiste una misura di errore. Per alcuni di questi algoritmi è possibile definire una funzione di <<energia>> dei parametri della rete, la quale fornisce una misura continua della convergenza della rete verso una soluzione. Altrimenti si possono utilizzare altri indici, come ad esempio una misura delle variazioni delle sinapsi, in quanto il raggiungimento di una soluzione di

solito corrisponde ad una situazione di stabilità sinaptica, o ad una misura di variazione dello stato di attivazione dei nodi della rete per ogni singolo pattern. Nei paragrafi seguenti approfondiremo l'apprendimento delle reti supervisionate, non tratteremo invece quelle auto-organizzate.

## 7.7 Architettura dei collegamenti

Una rete neurale artificiale è costituita da una serie di neuroni, variamente connessi fra di loro in strati successivi. Le unità che sono in contatto con l'ambiente esterno, e ricevono stimoli da esso, sono dette *unità di input*, quelle che forniscono il risultato dell'elaborazione dello stimolo ricevuto, anch'esse quindi in contatto con l'ambiente esterno, sono dette *unità di output*, mentre quelle unità che si trovano negli strati intermedi fra questi due sono dette *unità nascoste*: esse, infatti, sono in relazione solo con altre unità. Uno schema rappresentativo di queste reti è quello seguente:

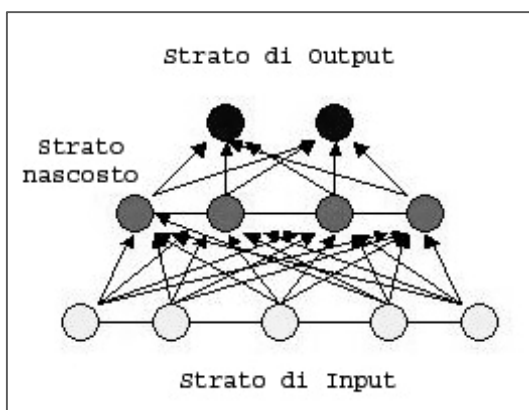


Figura 7.14 - Esempio di architettura dei collegamenti di una rete neurale.

L'architettura della rete neurale è determinata dal tipo di studio che si vuole realizzare ed è caratterizzata dalla distinzione tra neuroni di ingresso e di uscita, dal numero di strati di sinapsi (o neuroni) e dalla presenza di connessioni di retroazione.

## 7.8 Algoritmi di apprendimento

### 7.8.1 Regola di Hebb

Esistono diversi tipi di algoritmi di apprendimento che vengono scelti a seconda del compito che dovrà svolgere la rete neurale. Tutti gli algoritmi di apprendimento delle reti neurali hanno le loro fondamenta sulla regola di modifica sinaptica di Hebb e sulle sue variazioni.

La regola di Hebb dice che se due neuroni collegati fra loro sono contemporaneamente attivi, l'efficacia sinaptica della connessione viene rinforzata. Se consideriamo una rete neurale in cui i nodi di uscita possiedono una funzione di attivazione binaria, per ogni presentazione di un pattern di ingresso  $x$  e di un pattern di uscita  $y$ , la modifica dei pesi sinaptici sarà la seguente:

$$\Delta w_{ij} = \eta y_i x_j \quad [63]$$

dove  $\eta$  è il tasso di apprendimento. Se iniziamo la fase di apprendimento da una matrice di valori sinaptici  $w_{ij}$  uguale a 0 e fissiamo  $\eta = 1$ , la matrice finale sarà uguale alla somma di tutte le modifiche calcolate per ciascuna coppia di pattern  $\mu$

$$w_{ij} = \sum_{\mu} \Delta w_{ij}^{\mu} \quad [64]$$

Se presentassimo una versione incompleta di un pattern di ingresso la rete avrebbe le capacità di fornire ugualmente il dato in uscita corretto. Tuttavia, poiché la regola di Hebb prevede solamente l'incremento delle sinapsi, la rete non è in grado di apprendere associazioni che presentano elementi in comune ma che richiedono risposte diverse. Un pattern di input che presenta elementi in comune con un altro pattern di input causa l'attivazione dei nodi di output corrispondenti ad entrambi i pattern. La sovrapposizione delle coppie di *input* genera delle risposte di *output* dette “miste” e il fenomeno è chiamato di “interferenza”. La regola di Hebb permette di apprendere solamente pattern ortogonali per i quali la somma dei prodotti dei singoli componenti è zero.

### 7.8.2 Regola postsinaptica

La regola postsinaptica prevede che il valore della connessione sinaptica venga incrementato ogni volta che l'unità post sinaptica e quella presinaptica siano entrambe attive, e che venga invece diminuito ogni volta che l'unità postsinaptica è attiva mentre è inattiva l'unità presinaptica.

$$\Delta w_{ij} = \eta (y_i x_j + (x_j - 1) y_i) \quad [65]$$

### 7.8.3 Regola presinaptica

La regola presinaptica, rispetto alla precedente, prevede una situazione opposta in cui il valore della connessione sinaptica è aumentato quando entrambe le unità - presinaptica e postsinaptica - sono attive mentre viene indebolito quando solo l'unità presinaptica è attiva, come è indicato nella seguente relazione:

$$\Delta w_{ij} = \eta (y_i x_j + (y_j - 1) x_i) \quad [66]$$



### 7.8.4 Regola della covarianza

Conosciuta anche come regola di Hopfiel, la regola della covarianza è una combinazione delle precedenti. Con tale regola una connessione viene rinforzata quando entrambe le unità, presinaptica e postsinaptica, si trovano nello stesso stato, cioè entrambe attive o entrambe inattive; quando invece le due unità sono in stati diversi la connessione viene indebolita. Si evince però che: nel caso in cui la regola di attivazione è bipolare, la regola della covarianza assume la stessa forma della regola di Hebb divenendo capace di rafforzare solo le unità attive; in tal caso la regola della covarianza è definita come regola di “Hebb estesa”.

### 7.8.5 Regola di apprendimento del perceptrone semplice

Fu Rosenblatt il primo ad avere l'idea di modificare i pesi delle connessioni di una rete neurale artificiale in funzione dell'errore generato dal confronto tra la risposta della rete e la risposta desiderata. I perceptroni di Rosenblatt erano delle reti neurali con due strati di connessioni in cui le sinapsi del primo strato erano fissate in modo casuale e non potevano essere cambiate mentre le sinapsi del secondo strato erano sottoposte ad apprendimento. Per cui quando si parla di perceptroni ci si riferisce ad un unico strato di connessioni unidirezionali dai nodi di input ai nodi di output; una rete così realizzata cioè alimentata strettamente in avanti viene indicata con il termine *feed-forward*. I perceptroni sono detti semplici poiché i nodi di uscita utilizzano una semplice funzione a gradino. Data una rete neurale con N nodi d'ingresso e P coppie di addestramento, ciascuna composta da un vettore di ingresso bipolare  $x_p$  e da una risposta desiderata  $t$  anch'essa bipolare, l'output della rete per ciascun pattern d'ingresso è dato da:

$$y = \begin{cases} 1 & \text{se } \sum_{i=1}^N w_i x_i > b \\ -1 & \text{altrimenti} \end{cases} \quad [67]$$

Questo valore viene confrontato con la risposta desiderata  $t$  per quel determinato pattern di ingresso. Se le due risposte sono uguali non si avrà alcuna modifica dei valori sinaptici; se invece si è verificato un errore i pesi sinaptici subiranno delle modifiche proprio in base alla risposta corretta:

$$\Delta w_i = \begin{cases} 0 & \text{se } y = t \\ \eta y_d x_i & \text{altrimenti} \end{cases} \quad [68]$$

Questo valore viene dunque addizionato ai valori precedenti delle sinapsi:

$$w_i^t = w_i^{t-1} + \eta \Delta w_i^t \quad [69]$$

Le connessioni sinaitiche sono di solito prese con valori iniziali pari a zero o con piccoli valori presi casualmente o ancora possono essere uguali ad uno qualsiasi dei pattern di addestramento. Il tasso di apprendimento  $\eta$  viene di solito preso uguale ad 1. I percettoni modificano i loro pesi sinaptici solamente per i pattern di ingresso che provocano risposte scorrette e necessitano di molte presentazioni della lista di coppie di addestramento: man mano che l'addestramento procede le modifiche sinaitiche diventano sempre meno frequenti perché un numero via via minore di coppie di input provoca risposte non esatte. L'apprendimento si interrompe automaticamente quando nessun pattern provoca errori di risposta. In ogni caso se esiste un vettore di valori sinaptici che soddisfa la trasformazione richiesta da input ad output, la regola di apprendimento è in grado di trovare tale vettore in un numero finito di cicli di apprendimento. La condizione necessaria affinché vi sia una soluzione è che i pattern siano linearmente separabili.

### 7.8.6 Regola delta

La regola Delta è simile alla regola di apprendimento dei percettoni ed è applicabile ad unità di output dotate di una funzione di attivazione continua e differenziabile. Il vantaggio di questa regola è che permette di descrivere le prestazioni della rete neurale tramite una funzione continua  $E_w$  che misura l'errore della rete; questa misura è differenziabile rispetto alla matrice dei pesi sinaptici  $W$ , ovvero è possibile descrivere il cambiamento dell'errore in funzione del cambiamento dei pesi sinaptici.

Consideriamo una rete neurale di tipo feed-forward con unità di output ad attivazione lineare:

$$y_i = \sum_{j=0}^N w_{ij} x_j \quad [70]$$

Dato un gruppo di pattern di addestramento composto da  $M$  coppie formate dal vettore in ingresso  $x^\mu$  e dal vettore di risposta desiderata  $t^\mu$ , vogliamo trovare la matrice dei pesi sinaptici grazie alla quale si ottiene:

$$y_i^\mu = t_i^\mu \quad \forall i, \mu \quad [71]$$

Descriviamo le prestazioni generali della rete tramite la funzione di errore (detta anche funzione di costo)  $E_w$ :

$$E_w = \frac{1}{2} \sum_{\mu} \sum_i (t_i^\mu - y_i^\mu)^2 \quad [72]$$

nonché, tramite lo scarto quadratico medio tra la risposta desiderata e la risposta ottenuta per ciascuna unità della rete sommato su tutte le unità e su tutte le coppie di apprendimento. La funzione assume

valori tanto più piccoli quanto più le risposte della rete neurale si avvicinano alle risposte desiderate. Possiamo ancora scrivere:

$$E_W = \frac{1}{2} \sum_{\mu} \sum_i (t_i^{\mu} - \sum_{j=0}^N w_{ij} x_j^{\mu})^2 \quad [73]$$

Risulta evidente come la variazione della funzione  $E_W$  dipenda unicamente dal valore delle connessioni sinaptiche  $W$ . Infatti il gradiente (e quindi la direzione di massima crescita) della funzione è individuato da un vettore di derivate parziali di  $E_W$  rispetto a ciascun peso sinaptico  $w_{ij}$ . Poiché il nostro obiettivo è ridurre il valore di  $E_W$  e cioè il valore dello scarto quadratico medio tra la risposta desiderata e la risposta ottenuta, dobbiamo modificare i pesi sinaptici nella direzione opposta a quella del gradiente di  $E_W$ :

$$\Delta w_{ij} = - \frac{\partial E}{\partial w_{ij}} \quad [74]$$

Essendo ciascuna unità di output indipendente dalle altre e risolvendo la derivata per ciascuna singola unità, otteniamo:

$$\Delta w_{ij} = \sum_{\mu} (t_i^{\mu} - y_i^{\mu}) x_j^{\mu} \quad [75]$$

dunque, il cambiamento dei pesi sinaptici per ciascuna coppia di apprendimento  $\mu$  è dato dalla differenza tra la risposta desiderata e l'output della rete moltiplicata per l'attività presinaptica

$$\Delta w_{ij} = \eta (t_i - y_i) x_j \quad [76]$$

anche scritta come:

$$\Delta w_{ij} = \eta \delta_i x_j \quad [77]$$

dove  $\eta$  è il tasso di apprendimento già visto precedentemente.

Rispetto ai percettori semplici, la regola delta è in grado di operare con funzioni di attivazioni continue. La condizione per il raggiungimento di una soluzione è l'indipendenza lineare dei pattern di ingresso. Come già detto, il funzionamento della regola delta si basa sulla ricerca dei pesi sinaptici che corrispondono a valori minimi della funzione di errore  $E_W$ . Immaginiamo, allora, la funzione di errore come una superficie in uno spazio  $n$ -dimensionale (dove  $n$  è il numero dei pesi sinaptici della rete); essendo che ciascuna dimensione rappresenta lo spazio di possibile variazione di ciascuna sinapsi, la superficie presenta un aspetto variegato caratterizzato da <<minimi e massimi>>. Nel caso in cui i pattern di input siano linearmente indipendenti, la funzione di costo  $E_W$  per unità lineari presenta una superficie a forma di scodella con un minimo dove  $E_W=0$ . Se utilizziamo un tasso di apprendimento sufficientemente piccolo la modifica graduale dei pesi sinaptici corrisponde ad una discesa sulle pareti

della superficie verso il fondo (metodo di discesa del gradiente). Nel caso in cui i pattern di addestramento non siano linearmente indipendenti, la regola delta produce pur sempre una discesa sulla superficie della funzione di errore, ma il minimo raggiungibile può essere maggiore di zero; in altre parole la rete neurale trova solamente una soluzione parziale per i pattern di addestramento (ad esempio solamente alcune associazioni sono apprese correttamente oppure tutte le associazioni presentano qualche unità di output scorretta). Il tasso di apprendimento  $\eta$  controlla il movimento sulla superficie dell'errore: se  $\eta$  è troppo grande si rischia di passare da una parete all'altra senza raggiungere il minimo.

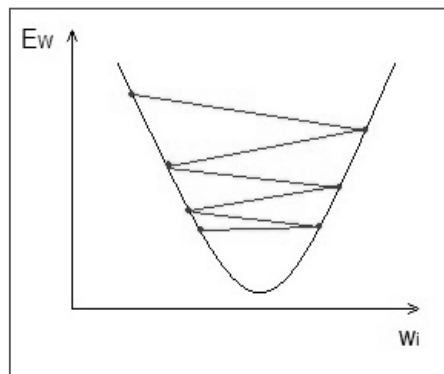


Figura 7.15 - Tasso di apprendimento troppo grande.

Se  $\eta$  è invece troppo piccolo la discesa verso il minimo è assicurata (solo nel caso di unità lineari), ma è estremamente lenta e richiede molte presentazioni delle coppie di addestramento.

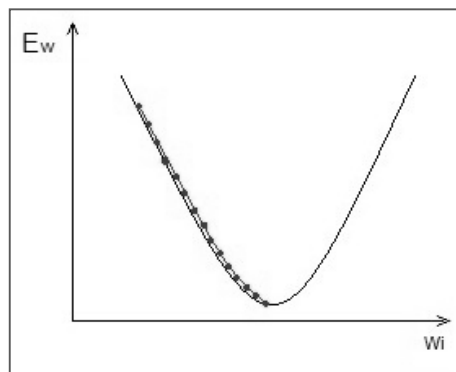


Figura 7.16 - Tasso di apprendimento troppo piccolo.

Alcuni autori hanno suggerito di scegliere il valore in base a  $0.1 \leq N\eta \leq 1,0$  dove  $N$  è il numero di unità di input per ogni unità di output (Widrow, Krogh e Palmer 1991); altri hanno indicato margini ben più ampi, come  $0,1 \leq \eta \leq 10,0$  (Hecht – Nielsen 1990).

Consideriamo adesso una rete neurale di tipo feed-forward in cui i nodi di uscita hanno funzione di attivazione continua e differenziabile non lineare  $y_i = \Phi(A_i)$ : in questo caso la funzione di costo assume la seguente espressione

$$E_W = \frac{1}{2} \sum_{\mu} \sum_i (t_i^{\mu} - \Phi(\sum_{j=0}^N w_{ij} x_j))^2 \quad [78]$$

Questa volta la discesa del gradiente della funzione rispetto ai pesi sinaptici produce la seguente regola di modifica sinaptica:

$$\Delta w_{ij} = -\frac{\partial E}{\partial w_{ij}} = \sum_{\mu} (t_i^{\mu} - y_i^{\mu}) \dot{\Phi}(A_i^{\mu}) x_j^{\mu} \quad [79]$$

quindi per ogni coppia di addestramento la modifica dei pesi sinaptici è

$$\Delta w_{ij} = \eta (t_i - y_i) \dot{\Phi}(A_i) x_j \quad [80]$$

Dove  $\dot{\Phi}(A_i)$  è la derivata prima della funzione di attivazione.

Ad esempio, se la funzione di attivazione è una sigmoidea tra 0 e 1,

$$y_i = \Phi(A_i) = (1 + e^{-\beta A_i})^{-1} \quad [81]$$

e la derivata è esprimibile direttamente in base all'output dell'unità stessa

$$\dot{\Phi}(A_i) = \beta y_i (1 - y_i). \quad [82]$$

L'impiego di unità continue non-lineari non modifica la condizione di raggiungimento di una soluzione, che consiste sempre nell'indipendenza lineare dei pattern di ingresso ma aumenta, rispetto al caso di funzioni di attivazione lineari, il numero di soluzioni parziali raggiungibili quando i pattern sono linearmente dipendenti. È da notare che con l'impiego di funzioni asintotiche (come la funzione sigmoide o la tangente iperbolica) la rete non è in grado di raggiungere il minimo assoluto  $E_W = 0$  quando le risposte desiderate si collocano agli estremi (0, 1 per la funzione sigmoide e +1, -1 per la tangente iperbolica) ma si porta su valori molto prossimi a zero. La superficie della funzione di errore con funzioni continue non-lineari può possedere diversi <<minimi locali>> assumendo una forma più complicata rispetto a quella a scodella vista nel caso di funzioni lineari, e può generare un errore relativamente piccolo ma che non corrisponde sempre alla risposta desiderata.

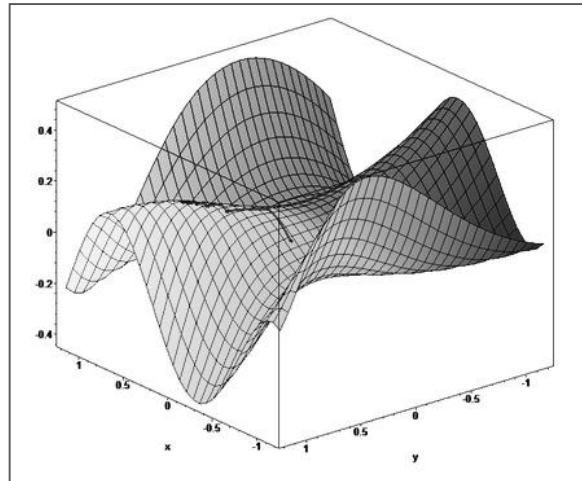


Figura 7.17 - Superficie della funzione di errore  $E_W$  per una rete multistrato con unità continue non-lineari.

### 7.8.7 Algoritmo di back-propagation e reti neurali multistrato

Definito anche con il nome di *regola delta generalizzata*, l'algoritmo di Back propagation è un'estensione della regola delta per reti neurali multi-strato; queste reti con unità ad attivazione non lineare hanno la capacità di classificare input che non sono linearmente separabili: esse infatti riescono a risolvere tutte le funzioni booleane, tra cui ad esempio lo  $X_{OR}$  (funzione molto semplice utilizzata come indice di prestazione di un algoritmo perché non linearmente separabile). Una rete *multi-layer* è organizzata in *layer* sequenziali, di cui il primo è quello dei dati di input e l'ultimo è quello dei parametri di output del modello. I *layer* intermedi, come già detto, sono chiamati "*hidden*", e possono essere uno o più di uno. Ogni *layer* è costituito da uno o più "neuroni" nonché da elementi che operano in parallelo o unità di elaborazione dei dati. Per l'*input layer*, ogni parametro fisico di cui si abbia a disposizione parametri sperimentali e che si suppone abbia influenza sul fenomeno sottoposto a modellazione, costituisce un neurone. Il *layer* di output contiene invece tanti neuroni quanti sono i parametri che ci si presuppone di ottenere con la rete neurale. Infine gli *hidden layer* contengono un numero variabile di neuroni; non esiste una regola generale per determinare il numero di neuroni negli *hidden layer*, di solito, la procedura è quella di testare diverse soluzioni e scegliere quella che fornisce le predizioni migliori. Nonostante già negli anni Sessanta si sapesse del vantaggio offerto da reti multistrato non-lineari, non esisteva ancora un algoritmo di apprendimento che fosse in grado di calcolare il contributo delle unità interne alla funzione di errore misurata sulle unità di output e quindi non esisteva un metodo per modificare le connessioni sinaptiche degli strati inferiori. Una soluzione a questo problema fu fornita dal metodo di Back-propagation, scoperto molteplici volte in modo indipendente e in diversi campi di ricerca. L'algoritmo di Back-propagation, come la regola delta, modifica i pesi sinaptici in base alla discrepanza tra la risposta fornita dalla rete e la risposta corretta. Prendiamo in considerazione una rete feed-forward a due strati di connessioni con unità di input  $\chi_k$ ,

unità interne “nascoste”  $h_j$  e unità di output  $y_i$ ; indichiamo il valore della risposta corretta per ogni unità di output con il simbolo  $t_i^\mu$ , dove  $\mu$  indica uno specifico pattern di ingresso. Ciascuna delle unità di output riceve connessioni sinaptiche  $w_{ij}$  da tutte le unità nascoste e ciascuna unità nascosta riceve connessioni sinaptiche  $v_{jk}$  da tutte le unità di input.

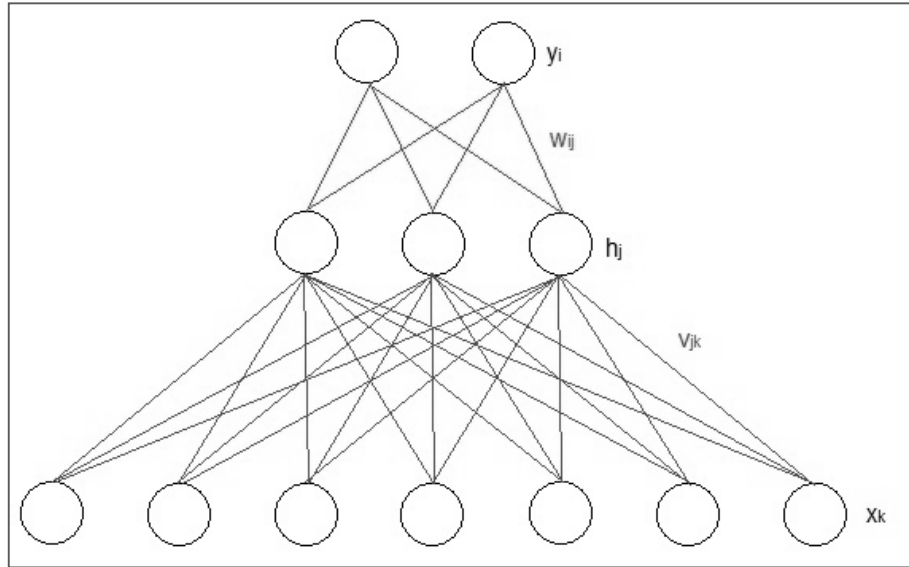


Figura 7.18 - Rete multistrato feed-forward con unità non lineari.

Tutte le unità, eccetto quelle di input, il cui valore corrisponde al pattern di ingresso, possiedono un'attivazione continua non-lineare, come ad esempio la funzione sigmoidea. Le unità di ingresso possono assumere valori binari, bipoli o continui. Per un dato pattern di ingresso  $\mathbf{x}^\mu$  (indicato in grassetto poiché è un vettore), si calcola innanzitutto l'attivazione di tutte le unità interne  $h_j$

$$h_j^\mu = \Phi(\sum_{k=0} v_{jk} x_k^\mu) \quad [83]$$

e successivamente l'attivazione di tutte le unità di output  $y_i$

$$y_i^\mu = \Phi(\sum_{j=0} w_{ij} h_j^\mu) \quad [84]$$

In entrambi i casi le unità possiedono una connessione (corrispondente all'indice 0) con un'unità di bias con valore costante pari a -1. Lo scopo è ridurre la differenza tra la risposta desiderata  $t_i^\mu$  e quella ottenuta  $y_i^\mu$ ; la funzione di errore  $E_w$  è la stessa vista per la regola delta

$$E_w = \frac{1}{2} \sum_{\mu} \sum_i (t_i^\mu - y_i^\mu)^2 \quad [85]$$

$$E_W = \frac{1}{2} \sum_{\mu} \sum_i (t_i^{\mu} - \Phi(\sum_j w_{ij} \Phi(\sum_k v_{ik} x_k^{\mu})))^2 \quad [86]$$

Questa volta la funzione di errore dipende da due gruppi di connessioni sinaptiche ( $w_{ij}$  e  $v_{jk}$ ); per ottenere la regola di modifica sinaptica è sempre possibile utilizzare il metodo di discesa del gradiente della funzione. Per la matrice di connessioni  $w_{ij}$  tra le unità nascoste e le unità di output abbiamo lo stesso risultato già visto per la regola delta

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = \eta \sum_{\mu} (t_i^{\mu} - y_i^{\mu}) \dot{\Phi}(A_i^{\mu}) h_j^{\mu} \quad [87]$$

che per semplicità possiamo scrivere come

$$\Delta w_{ij} = \eta \sum_{\mu} \delta_i^{\mu} h_j^{\mu} \quad [88]$$

dove

$$\delta_i^{\mu} = (t_i^{\mu} - y_i^{\mu}) \dot{\Phi}(A_i^{\mu}) \quad [89]$$

Per calcolare la modifica delle connessioni dello strato inferiore  $v_{jk}$  è stata utilizzata la regola di derivazione delle funzioni composte ed è stata considerata la variazione dell'attivazione delle unità interne rispetto alla variazione dei pesi sinaptici  $v_{jk}$  che trasmettono segnali da un input costante.

$$\Delta v_{jk} = -\eta \frac{\partial E}{\partial v_{jk}} = \eta \sum_{\mu} \frac{\partial E}{\partial h_j^{\mu}} \frac{\partial h_j^{\mu}}{\partial v_{jk}} \quad [90]$$

che si sviluppa nel modo seguente

$$\Delta v_{jk} = \eta \sum_{\mu} \sum_i (t_i^{\mu} - y_i^{\mu}) \dot{\Phi}(A_i^{\mu}) w_{ij} \dot{\Phi}(A_j^{\mu}) x_k^{\mu} \quad [91]$$

che, ricordando come abbiamo definito  $\delta_i^{\mu}$ , possiamo ancora scrivere come

$$\Delta v_{jk} = \eta \sum_{\mu} \sum_i \delta_i^{\mu} w_{ij} \dot{\Phi}(A_j^{\mu}) x_k^{\mu} \quad [92]$$

È importante notare che i pesi sinaptici dello strato inferiore richiedono per ciascuna unità interna  $h_j$  la somma dei prodotti tra i delta  $\delta_i^{\mu}$  dello strato superiore di unità a cui l'unità nascosta è collegata e i corrispondenti pesi sinaptici  $w_{ij}$ . In altre parole gli errori (i delta) dello strato superiore vengono



*propagati all'indietro* (da cui il nome di Back-propagation) attraverso le stesse connessioni sinaptiche e sommati per ciascuna unità inferiore da cui ricevono il segnale. Se poniamo i delta per le unità nascoste pari a

$$\delta_j^\mu = \phi(A_j^\mu) \sum_i w_{ij} \delta_i^\mu \quad [93]$$

la modifica dei pesi sinaptici  $v_{jk}$  assume la stessa forma della modifica dei pesi sinaptici  $w_{ij}$

$$\Delta v_{jk} = \eta \sum_\mu \delta_j^\mu x_k^\mu \quad [94]$$

Grazie al fatto che i delta di uno strato di unità sono ottenuti in base ai delta dello strato superiore, la regola di Back-propagation si applica a reti neurali con numero qualsiasi di strati, ciascuno composto da un numero qualsiasi di nodi.

Dunque, il funzionamento dell'algoritmo consiste nel presentare i pattern di input e propagare l'attivazione strato per strato fino alle unità di uscita. Il successo dell'algoritmo di Back-propagation è dovuto al fatto che, come già detto, permette di risolvere problemi non linearmente separabili ed ha un funzionamento semplice e consente di utilizzare reti con architetture di qualsiasi tipo. Differentemente dalla regola delta, la funzione di costo  $E_w$  può assumere andamenti più complessi dovuti alla sequenza di operazioni non lineari eseguiti durante il calcolo dell'attivazione e spesso presenta molti minimi locali (soluzioni parziali), zone caratterizzate da ampie superfici piatte, improvvisi dirupi, etc.. I minimi locali corrispondono a delle regioni in cui l'algoritmo può rimanere intrappolato senza poter raggiungere la soluzione; le zone piatte invece sono pericolose per la convergenza perché in queste situazioni la derivata prima dell'attivazione del nodo postsinaptico è prossima a zero e i pesi sinaptici vengono quindi modificati pochissimo. Nonostante le varie difficoltà che può incontrare, l'algoritmo di Back-propagation è molto efficace nel raggiungere soluzioni soddisfacenti e a generalizzare a nuovi esempi che non sono stati presentati in fase di addestramento.

## **7.9 Problemi nella generalizzazione**

I problemi legati ad una adeguata generalizzazione si presentano quando possediamo un numero limitato di pattern di addestramento ma vogliamo avere un buon livello di prestazioni quando, completato l'addestramento, si presentano dei nuovi pattern. Le reti multistrato che possiedono un gran numero di strati nascosti, e quindi un gran numero di risorse (pesi e unità), possono apprendere una vasta serie di funzioni specifiche che sono responsabili della corrispondenza tra *input* e *output* per i pattern di addestramento, diminuendo così la probabilità che la rete scopra la funzione generale che descrive l'intero dominio del problema e che darebbe luogo a risposte corrette anche per i pattern di

test. Per richiamare le approssimazioni in statistica, se il numero di parametri liberi (nel nostro caso i pesi della rete) da determinare è troppo elevato, si ottiene una descrizione esatta dei dati ma si rischiano errori per i nuovi dati.

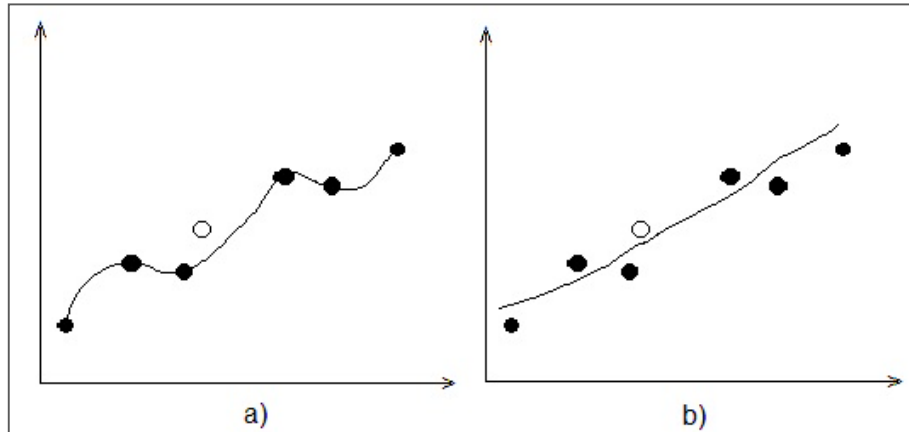


Figura 7.19 - Apprendimento della funzione che descrive i dati di addestramento (cerchietti neri). L'errore è lo scarto quadratico medio tra un dato e il corrispondente punto della funzione. a) Un numero troppo grande di parametri non garantisce una buona generalizzazione del nuovo pattern (cerchietto bianco); b) un numero adeguato di parametri permette una migliore generalizzazione (lo scarto quadratico medio è inferiore).

Un numero troppo grande di connessioni compromette anche la convergenza verso un minimo globale, sia perché aumenta la complessità della superficie dell'errore, sia perché richiede un alto numero di pattern di addestramento. Un modo per limitare il problema della scarsa generalizzazione, evitando il problema dell'architettura della rete, è quello di interrompere l'addestramento prima che la rete apprenda <<troppo bene>> i pattern di addestramento, cioè prima che si verifichi il fenomeno dell'*overfitting*.

In questo modo si tenta di riprodurre una situazione analoga a quella di Figura 7.19 b), in cui la rete, benché non raggiunga un minimo globale per i pattern di addestramento, è comunque capace di raggiungere un livello accettabile di prestazioni sia sui pattern di addestramento che su quelli di test. Quello che viene molto spesso fatto è che si divide in tre parti, non necessariamente uguali, la collezione dei pattern disponibili; una parte per l'addestramento della rete, un'altra per la <<validazione>> dell'apprendimento e un'altra ancora di test per valutare la capacità di generalizzazione. Mentre la modifica dei pesi sinaptici si basa sull'errore calcolato per i pattern di addestramento, la decisione su quando fermare l'apprendimento si basa sull'errore ottenuto per i pattern di validazione.

È possibile verificare che l'errore sui pattern di validazione e di test inizialmente si riduce in modo simile all'errore sui pattern di addestramento, ma ad un certo punto l'errore sui pattern di validazione e di test comincia a crescere mentre l'errore sui pattern di addestramento continua a diminuire. Questa inversione di tendenza indica l'inizio del <<sovrapprendimento>> (o *overfitting*) dei dati di addestramento e rappresenta un buon indicatore di quando fermare l'apprendimento. La decisione si basa sulla registrazione dell'errore per il gruppo di validazione nelle ultime  $N$  epoche di addestramento: se esso non diminuisce per tutte le  $N$  epoche, si cessa l'addestramento e si mantengono i pesi sinaptici corrispondenti all'errore minore nelle  $N$  epoche.

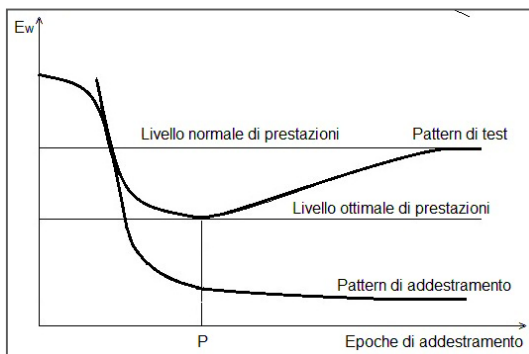


Figura 7.20 - Andamento dell'errore sui pattern di addestramento e di test (più i pattern di validazione); il livello ottimale di prestazioni si ottiene se l'apprendimento viene fermato prima che si verifichi un overfit dei dati di addestramento (punto P).

## 7.10 Il modello GMDH e le reti neurali

Nella fase iniziale dello sviluppo della teoria GMDH erano abbastanza evidenti le somiglianze tra le Neural Network e gli algoritmi multistrato GMDH. Tanto che più articoli scientifici hanno sostenuto che, poiché le differenze tra perceptron e GMDH sono né significative né fondamentale è opportuno chiamare sistemi GMDH come “sistemi di tipo perceptron”.

In primo luogo le Neural Network sono state inserite nel modello GMDH per la determinazione delle regole Fuzzy dello stesso algoritmo. Tuttavia, l'applicazione delle Neural Network ha prodotto non pochi problemi di ottimizzazione, che sono stati successivamente eliminati con un algoritmo multistrato GMDH che permette di fatto che non si abbia un numero elevato di regole generate.

Un certo numero di ricercatori giapponesi nell'ultimo decennio, motivato dalle somiglianze tra le Neural Network e GMDH, hanno progettato una serie di algoritmi multistrato che sono stati introdotti come Neural network GMDH. La caratteristica principale di questi algoritmi è l'adozione di criteri interni per la selezione del processo di apprendimento, che possono non richiedere la partizione del set di dati. Tale approccio ha però subito delle critiche basate sul fatto che una delle caratteristiche

principali del modello GMDH è l'applicazione del principio del criterio esterno e quindi ogni algoritmo che non soddisfa questo principio non dovrebbe essere inserito tra i modelli GMDH. Tuttavia, l'*Akaike Information Criterion* (AIC), introdotto con le Neural Network, nonché il criterio di polarizzazione minima differenziale (DMC), hanno dato buoni risultati nel selezionare un modello ottimale e quindi il numero di strati nascosti della Neural network GMDH.

Pertanto le analogie tra la struttura delle Neural Network e algoritmi GMDH hanno spinto i ricercatori a esplorare le zone in cui questi due concetti potrebbero combinarsi. Partendo dal fatto che metodi induttivi non possono sostituire le analisi teoriche necessarie sulle cause degli eventi, una pragmatica soluzione ai problemi di costruzione dei modelli può essere affrontata con l'unione delle metodologie deduttive e induttive, e quindi con un metodo combinato, che estende la teoria di auto-organizzazione da modelli isolati a reti neurali attive.

L'algoritmo proposto è noto come "reti neurali con neuroni attivi", dove al posto del passivo vengono applicati gli algoritmi Neural network GMDH. Le Neural Network con neuroni attivi hanno una duplice struttura multistrato, infatti i neuroni stessi sono multistrato e possono essere uniti in comuni matrici in modalità multilayered. Pertanto si sono utilizzati gli algoritmi multistrato e combinatori come neuroni attivi al fine di aumentare la precisione e ridurre i tempi di calcolo.

Una Neural Network con neuroni attivi ha quindi la capacità di selezionare gli argomenti di input più rilevanti così da aumentare la sua precisione nei compiti di previsione, superando l'imperscrutabilità delle reti neurali con neuroni passivi. Inoltre, la precisione viene aumentata poiché l'uscita di neuroni attivi genera nuove variabili, che possono essere utilizzate come un elemento nuovo per il successivo strato di neuroni e nel mentre i fattori di input saranno ottimizzati in ogni layer.

### **7.10.1 GMDH type Neural Networks**

GMDH tipo Neural network applica delle reti neurali polinomiali attraverso l'utilizzo di algoritmi combinatori per l'ottimizzazione dei collegamenti tra i neuroni. L'algoritmo genera iterativamente strati di neuroni con due o più ingressi dei quali solo un numero limitato definisce la larghezza iniziale del layer. Ogni nuovo livello viene creato utilizzando due o più neuroni prelevati da uno qualsiasi dei layer precedenti. Ogni neurone nella rete applica più funzioni di trasferimento, di solito con due variabili, al fine di realizzare una esaustiva ricerca volta a determinare la funzione di trasferimento migliore. Generalmente, la funzione di trasferimento ha una forma quadratica o lineare, ma può essere anche personalizzata. Come abbiamo già detto, l'algoritmo restituisce solo un numero limitato di neuroni definiti per ogni strato. Dal momento che ogni nuovo livello è in grado di connettersi ai livelli precedenti la larghezza del layer cresce costantemente. Poiché risulta raro che strati superiori tendono a migliorare il modello, solo metà dei neuroni generati al livello precedente sarà presente nel layer successivo, cioè il numero di neuroni  $N$  a livello  $k$   $N_k = 0,5 \cdot N_{k-1}$ . Questa assunzione rende l'algoritmo più veloce, mentre la possibilità di ridurre la qualità del modello è basso.

La generazione di nuovi layer è quindi arrestata quando un nuovo livello non fornisce una maggiore precisione nei test rispetto al livello precedente. Tuttavia, potrà essere possibile arrestare la generazione di nuovi livelli quando l'errore di test è stato ridotto a meno dell'1% o se il numero di strati ha raggiunto un certo limite definito.

## CAPITOLO 8 ***I Bilanci energetici locali***

La soluzione dei molteplici problemi posti da una attività di programmazione energetica, non può trascendere da una conoscenza approfondita del sistema energetico della realtà territoriale indagata, nelle varie sfaccettature disegnate dalle particolarità locali; chiunque abbia intenzione di dare risposta alle pressanti esigenze poste dalla società e dall'economia in termini energetici e, conseguentemente, ambientali, non può agire senza la conoscenza delle variabili quantitative su cui andrà ad operare e con cui dovrà interagire [34]. Una programmazione efficace ed efficiente, responsabile da un punto di vista economico e sociale, che abbia come primario interesse il progresso del territorio amministrato, deve avere una chiara visione dell'esistente, così come del passato, più o meno recente, in modo da indagare e prevedere quale sarà il futuro, agendo così non di "rimessa", tamponando ciò che è già accaduto, ma proponendo attivamente interventi in grado di indirizzare il territorio sul percorso ritenuto più equilibrato. Per fare questo, il "programmatore" ha la necessità di uno strumento che gli consenta di ottenere tale visione globale dei fenomeni interessati e tale strumento è il Bilancio Energetico Locale (B.E.L.).

La disponibilità dei Bilanci Energetici è propedeutica alla stesura e all'elaborazione dei Piani Energetico Ambientali a livello territoriale. In questo contesto, il B.E.L. esplica la sua insostituibile funzione conoscitiva estrinsecando in un quadro riepilogativo "quanta" e che "tipo" di energia è stata consumata in un dato periodo di tempo (ad esempio un anno) e "come" essa è stata "prodotta", "reperita" sui mercati, "trasformata" e "consumata", all'interno della regione oggetto dello studio offrendo un'immagine immediata e sintetica del settore energetico di una data area.

Il B.E.L. offre quindi un quadro di sintesi che permette:

- di seguire l'evoluzione della domanda e dell'offerta di energia attraverso il confronto tra bilanci energetici relativi a diversi esercizi;
- di fare un confronto con la situazione energetica nazionale evidenziandone diversità e problemi;
- di valutare le interrelazioni con il sistema socio-economico.

In particolare, il confronto tra i bilanci energetici relativi a diversi periodi di tempo permette di seguire l'evoluzione del settore energetico, le sue interrelazioni con la struttura socio-economica e gli effetti di interventi tesi alla sua razionalizzazione, al fine di impostare la "programmazione energetica" a livello nazionale e/o regionale.

La redazione dei B.E.L. porta con sé notevoli difficoltà di compilazione, in quanto per assicurare la sua completa intellegibilità e confrontabilità, dovrebbe rispondere a criteri di uniformità e contare su dati completi e standardizzati, come in realtà non è, a causa dell'insufficienza o dell'assenza di rilevazioni uniformi e tipizzate, e dalla difficoltà a reperire dati sufficientemente disaggregati od attendibili.

Infatti, le principali difficoltà riguardano:

- il ritardo con cui i dati energetici sono resi disponibili;
- la scarsa disaggregazione che viene fatta di questi dati a livello territoriale oltre che settoriale;
- l'elevata frammentarietà dei dati, raccolti con criteri di classificazione che differiscono da ente ad ente, che portano spesso ad avere dati completamente diversi per uno stesso fenomeno energetico;
- la difficoltà ad accedere ad alcuni dati, soprattutto di aziende petrolifere.

La redazione del bilancio energetico presenta quindi non pochi problemi, sia nel reperimento del materiale statistico e delle serie storiche, sia nell'adozione di particolari convenzioni per ovviare all'assenza dei primi, in modo da stimare coerentemente quantità sconosciute o solo parzialmente osservabili. Le difficoltà non si limitano tuttavia solo ai dati: si estendono anche alla conversione del contributo energetico delle diverse fonti in una misura comune e facilmente confrontabile, così da rendere uniforme e facilmente interpretabile il contenuto del bilancio stesso. Per aggregare i dati quantitativi delle varie fonti energetiche si fa ricorso ad una operazione di conversione attraverso la quale le unità di misura delle varie fonti energetiche sono sostituite con una unità comune che permette la loro aggregazione a livello globale. Le unità più comunemente utilizzate per elaborare bilanci energetici sono: la caloria, la tonnellata equivalente di carbone (tec) e la tonnellata equivalente di petrolio (tep).

## **8.1 Il Bilancio Energetico locale: aspetti metodologici**

Il B.E.L. è costituito, da un modello di contabilità energetica che descrive la formazione delle disponibilità (offerta di energia) e degli impieghi (domanda) di fonti energetiche che si realizza in un dato periodo di tempo nel sistema economico e sociale osservato. Il patrimonio informativo che un B.E.L. fornisce, permette di effettuare due tipologie di analisi, note comunemente in letteratura come *analisi strutturali* e *analisi di impatto* o *previsionali*.

Le tecniche di analisi strutturale vanno dalla semplice illustrazione delle equazioni utilizzate per la redazione del B.E.L., al calcolo di rapporti o di altri parametri caratteristici, che vengono impiegati per

costruire graduatorie dei settori produttivi o per effettuare confronti territoriali o temporali utili ad evidenziare gli aspetti specifici dell'economia osservata.

Nelle analisi di impatto, questo modello, si presta ad essere utilizzato per valutare l'effetto prodotto da manovre di politica economica, che operano facendo variare direttamente le componenti dei consumi finali energetici (un programma di investimenti per introdurre tecnologie energy-savings, per esempio), o per effettuare esercizi di simulazione a scopo predittivo.

Il B.E.L. per rispondere alle esigenze conoscitive e per essere uno strumento utile per chi deve prendere decisioni e operare nel settore dell'energia deve essere:

- disponibile tempestivamente;
- elaborabile con facilità;
- disaggregato in misura sufficiente per fornire un quadro significativo della situazione energetica regionale e della sua dinamica;
- integrabile e confrontabile con altri bilanci energetici regionali e/o con il Bilancio Energetico Nazionale così da costituire anche uno strumento di base conoscitivo per aree interregionali.

Prima di descrivere la struttura del B.E.L., occorre preliminarmente operare alcune scelte che condizionano la significatività dello stesso. Queste scelte riguardano:

- la struttura e l'articolazione delle poste che concorrono a costituire le disponibilità e gli impieghi;
- le fonti energetiche da contabilizzare;
- i coefficienti da adottare per rendere "omogenee" e quindi sommabili fra di loro le diverse fonti di energia.

In ogni caso il B.E.L. è compilato nel rispetto delle equivalenze tra l'energia immessa e l'energia ricavata, quest'ultima integrata con le perdite e i consumi avvenuti nella fase di produzione, trasformazione, trasporto, distribuzione ed utilizzo della stessa; cercando, inoltre, di rimanere il più possibile aderente ai propri obiettivi fondamentali, che nell'ambito della programmazione energetica sono strettamente legati alla struttura dei legami sottesi alla formazione della domanda e dell'offerta di energia e possono riassumersi in:

- *obiettivo conoscitivo*: sotto il profilo quantitativo e di individuazione dei flussi energetici in trasformazione ed in trasferimento;
- *obiettivo interpretativo*: consentendo lo studio delle correlazioni esistenti tra le variabili energetiche e socio-economiche tali da documentare eventuali analisi e supportare scelte di intervento.



Tra i due obiettivi si può individuare un “verso” di utilizzazione, essendo il primo funzionale al secondo nel supportare le successive decisioni politiche che, per trovare ambito di applicazione a livello territoriale, le autorità responsabili hanno la necessità di poggiare sulla più ampia disaggregazione delle poste, in modo da poter analizzare in maggior dettaglio i vari settori in entrata e di impiego finale.

## **8.2 La struttura generale del B.E.L.**

Da un punto di vista pratico, il bilancio energetico locale è costituito da una matrice composta da tre sezioni:

- La prima sezione (in cui è riportata l’offerta delle fonti energetiche primarie e derivate) evidenzia la disponibilità di fonti energetiche per il territorio considerato.
- La seconda sezione è costituita dal sistema della trasformazione delle fonti primarie in prodotti energetici; qui si computano le quantità di fonti in ingresso, le perdite di trasformazione, i consumi dei processi e le uscite dei prodotti finali destinati al consumo.
- La terza sezione è costituita dal sistema dei consumi finali; qui confluiscono tutte le forme di prodotti energetici (primarie e derivate) che vanno ad essere impiegate nei settori produttivi, residenziale, terziario e trasporti.

Nella matrice dei consumi finali appare anche una sottosezione in cui vengono separati gli impieghi non energetici delle fonti di energia, ovvero gli impieghi in qualità di materie prime che entrano nella produzione. La matrice comprendente il B.E.L. nel suo complesso è strutturata in modo che tutta l’informazione disponibile, sia memorizzata in una matrice  $A$  ( $62 \times 39$ ) all’interno della quale vengono costruite alcune sottomatrici rettangolari riferite alle seguenti classi omogenee di fonti energetiche:

- solidi;
- liquidi;
- gassosi;
- energia elettrica.

Nel B.E.L. viene confrontata per ogni fonte energetica la rispettiva disponibilità (poste attive) con i relativi impieghi (poste passive), in modo tale da ottenere l’identità fondamentale di un bilancio energetico caratterizzata dall’uguaglianza tra disponibilità ed impieghi.

Le poste attive (disponibilità) sono individuate da:

- produzione;
- saldo in entrata;
- saldo in uscita;
- variazione delle scorte a livello primario.

Le poste passive (impieghi) sono individuate da:

- trasformazioni;
- bunkeraggi;
- consumi e perdite del settore energetico;
- usi non energetici (consumi finali non energetici);
- usi energetici (consumi finali energetici).

Si esamina ora la struttura del B.E.L. dal lato delle “righe” e delle “colonne” delle matrici. Una rappresentazione compatta è rappresentata nella Figura 8.1.

|                                    | <b>Solidi</b> | <b>Petrolio</b> | <b>Gas</b> | <b>Rinnovabili</b> | <b>En. Elett.</b> | <b>Totale</b> |
|------------------------------------|---------------|-----------------|------------|--------------------|-------------------|---------------|
| 1) Produzione                      |               |                 |            |                    |                   |               |
| 2) Saldo in entrata                |               |                 |            |                    |                   |               |
| 3) Saldo in uscita                 |               |                 |            |                    |                   |               |
| 4) Var. delle scorte               |               |                 |            |                    |                   |               |
| <b>5) Cons. interno lordo</b>      |               |                 |            |                    |                   |               |
| 6) Cons. e perdite settore energia |               |                 |            |                    |                   |               |
| 7) Trasf. in en. elett.            |               |                 |            |                    |                   |               |
| 8) Bunkeraggi                      |               |                 |            |                    |                   |               |
| 9) Usi non energ.                  |               |                 |            |                    |                   |               |
| <b>10) Tot. Impieghi finali</b>    |               |                 |            |                    |                   |               |
| 11) Agric. e Pesca                 |               |                 |            |                    |                   |               |
| 12) Industria                      |               |                 |            |                    |                   |               |
| 13) Civile                         |               |                 |            |                    |                   |               |
| 13.1 Domestico                     |               |                 |            |                    |                   |               |
| 14) Trasporti                      |               |                 |            |                    |                   |               |

Figura 8.1 - Rappresentazione compatta del B.E.L..

### 8.2.1 Le righe

Le righe di ciascuna sottomatrice rappresentano le voci della formazione delle disponibilità (o risorse), delle trasformazioni e degli impieghi finali. Gli aggregati presenti in ogni sottomatrice rappresentano le voci logiche generali (poste attive e poste passive) organizzate secondo un modello gerarchico con “relazione di padre-figlio”.

Lo schema di bilancio è costituito da alcune parti principali:

### 1. Risorse

Questo aggregato costituisce il perno del bilancio e rappresenta, la quantità di energia disponibile per i consumi energetici e non energetici all'interno del territorio. La sezione delle risorse si compone di tre righe:

- *Produzione*: questa voce indica la quantità di energia effettivamente prodotta sul territorio regionale; comprende sia la produzione primaria che quella secondaria.
- *Saldo in entrata*: questa voce indica l'acquisto da altre regioni o da paesi esteri di fonti energetiche e la loro introduzione nel territorio regionale ad esclusione dei transiti, in particolare per gasdotto e oleodotto.
- *Variazioni delle scorte*: questa voce indica la differenza tra le quantità di fonti di energia esistenti presso il sistema primario all'inizio e alla fine del periodo considerato. Il segno positivo (+) indica un prelevamento dalle scorte e dunque un aumento delle risorse; il segno negativo (-) una costituzione di scorte e quindi una diminuzione delle risorse.

### 2. Trasformazioni

Questo aggregato comprende le unità produttive che attuano la produzione o la trasformazione di fonti di energia. Questa parte assicura il collegamento tra la parte “risorse” e la parte “impieghi”. L'attività di trasformazione si compone di tre righe:

- *Ingressi*: Indica i quantitativi di fonti energetiche primarie e/o secondarie che entrano (input) nei diversi impianti di trasformazione per ottenere fonti energetiche derivate (secondarie).
- *Perdite*: indica le perdite di fonti energetiche insite nel tipo di processo tecnologico di trasformazione utilizzato.
- *Uscite*: indica il risultato del processo di trasformazione e corrisponde alla produzione di prodotti derivati.

### 3. Consumi e perdite del settore energia

Questo aggregato indica i consumi propri di fonti di energia dovuti al funzionamento degli impianti di trasformazione o di autoproduzione e alle perdite di trasporto e distribuzione all'utente finale. In tale aggregato compaiono consumi di energia per:

- *Produzione e distribuzione di energia elettrica* (per il funzionamento dei servizi ausiliari delle centrali elettriche);

- *Perdite sulle reti:* (dovute al trasporto e alla distribuzione dell'energia elettrica e del gas naturale);
- *Produzione di energia elettrica* (saldo di pompaggio). Le perdite di pompaggio, cioè il saldo fra l'energia elettrica assorbita dal pompaggio e quella prodotta in seguito al pompaggio, vengono considerate come un consumo proprio dell'impianto di produzione e non come un'attività di trasformazione dell'energia elettrica (poiché la natura del prodotto non viene modificata).
- *Cokerie e officine del gas* (consumo per il funzionamento degli impianti delle cokerie e delle officine del gas);
- *Estrazione di petrolio e gas naturale* (consumo per il funzionamento degli impianti di estrazione);
- *Oleodotti e gasdotti* (quantitativi consumati nelle stazioni di compressione e pompaggio degli oleodotti e gasdotti);
- *Raffinerie di petrolio.*

#### 4. Non trasformate.

Questa voce indica i quantitativi di fonti di energia che non entrano nel processo di trasformazione.

#### 5. Totale disponibile.

Questa voce indica le quantità di fonti energetiche utilizzabili in ambito territoriale e corrisponde alla somma dei consumi interni (energetici e non energetici), dell'autoconsumo e perdite del settore energetico, delle esportazioni e dei bunkeraggi internazionali.

#### 6. Bunkeraggi.

Questa voce indica i rifornimenti (marittimi e aerei) di fonti energetiche fatti ad operatori esteri in ambito territoriale. I bunkeraggi sono assimilati al consumo originato dalla presenza di determinate strutture produttive e sono inclusi nel settore dei trasporti.

#### 7. Saldo in uscita.

Questa voce indica le fonti energetiche in uscita dal territorio (vedere quanto descritto alla voce "Saldo in entrata").

#### 8. Disponibilità interna.

Questa voce indica la quantità di fonti di energia messa a disposizione dell'utente finale. Tale disponibilità risulta dalla somma degli usi non energetici e degli usi energetici.

#### 9. Usi non energetici.

Questo aggregato indica le quantità di fonti energetiche, utilizzate come materia prima nei processi industriali nei settori della Chimica, Petrochimica ed altre branche di consumo a fini non energetici.

#### 10. Usi energetici (Consumo finale energetico)

Questo aggregato indica l'energia fornita all'utente finale per tutti gli impieghi energetici. A questo proposito, si distinguono i consumi finali di fonti energetiche nei quattro macrosettori Agricoltura e Pesca, Industria, Civile e Trasporti, per ciascuno dei quali, come evidenziato nella tabella che segue, si considera una suddivisione in branche.

### 8.2.2 *Le colonne*

Le colonne di ciascuna sottomatrice energetica individuano le fonti energetiche che vengono prese in considerazione per la costruzione del B.E.L.. Si fa riferimento sia alle fonti naturali di energia primaria, quali l'energia idrica, l'energia geotermica, i combustibili solidi, gli idrocarburi liquidi e gassosi, i combustibili nucleari, sia a quelle secondarie ottenute per trasformazione delle fonti primarie, come il gas d'officina, di cokeria e d'altoforno, i prodotti di raffinazione del petrolio, il coke da cokeria e da gas, il carbone di legna e l'energia elettrica. La rilevazione statistica viene applicata alle fonti energetiche che hanno una particolare rilevanza economica in quanto oggetto di scambi e/o autoconsumi sostitutivi degli acquisti sul mercato.

Le sottomatrici che compongono il B.E.L. sono riferite alle classi omogenee di fonti energetiche (solidi, liquidi, gassosi, energia elettrica, fonti rinnovabili). Ogni sottomatrice riporta nelle colonne la disaggregazione delle voci che compongono la classe omogenea.

Nel modello della figura 8.1 ogni fonte aggregata comprende sia vettori energetici primari che secondari. In particolare :

- *la voce “solidi” comprende:* carbone fossile, lignite, coke di cokeria, gas di cokeria, gas d'altoforno e altri prodotti solidi
- *la voce “petrolio” comprende:* il petrolio greggio, olio combustibile, gasolio, distillati leggeri, benzine, carboturbo, petrolio da riscaldamento, G.P.L., gas di raffineria e altri prodotti petroliferi;
- *la voce “gas”:* comprende il gas naturale e il gas d'officina;
- *la voce “rinnovabili”:* comprende la legna, il carbone da legna, le biomasse, e l'energia elettrica prodotta da fonte idraulica, eolica, solare, nucleare e geotermica;
- *la voce “energia elettrica”:* comprende il saldo in entrata ed in uscita e l'energia elettrica all'utenza finale.

## CAPITOLO 9 *I software utilizzati*

Come espresso nel capitolo 2, sulla base degli studi svolti, degli approfondimenti condotti e dei dati raccolti, sono stati identificati ed approfonditi i software che meglio si prestano agli scopi del lavoro.

La ricerca ha quindi condotto ad utilizzare:

- *il pacchetto IBM SPSS*, che è un software di analisi, classificazione e clustering di dati sperimentali;
- *il pacchetto GMDH Shell* (Forecasting Software for Business), che è un software di analisi, classificazione e clustering di dati sperimentali che realizza attraverso algoritmi complessi o Neural Network delle analisi previsionali.

In riferimento al software IBM SPSS di seguito saranno esposte le tematiche di analisi del dato, poiché per questo è stato utilizzato. Per quanto invece concerne il software GMDH Shell verranno esposte le modalità di clustering e forecasting.

### **9.1 Pacchetto IBM SPSS**

IBM® SPSS® Statistics è un sistema completo per l'analisi dei dati. *SPSS Statistics* è in grado di elaborare dati provenienti da diversi tipi di estensioni file e di utilizzarli per generare rapporti con tabelle, grafici, grafici di distribuzioni e trend, statistiche descrittive e complesse analisi statistiche. L'interfaccia grafica facilita l'utilizzo del software che tuttavia permette l'utilizzo del linguaggio a comandi per, in alcuni casi, l'applicazione di caratteristiche avanzate [35].

La piattaforma fornisce un'ampia gamma di procedure statistiche per analisi e report di base, inclusi conteggi, tavole di contingenza e statistiche descrittive, cubi OLAP e report con informazioni sui dati. Offre inoltre un'ampia varietà di tecniche di riduzione, classificazione e segmentazione di dimensioni, quali l'analisi fattoriale, l'analisi cluster, l'analisi del vicino più vicino e l'analisi della funzione discriminante. In più, propone un'ampia gamma di algoritmi per il confronto delle medie, nonché tecniche predittive quali il *test t*, l'analisi della varianza, la regressione lineare e la regressione ordinale.

Il programma è stato quindi utilizzato per l'analisi descrittiva e di frequenza dei dati.

### 9.1.1 Preparazione dei dati

Caricato il file DataSet, è possibile creare report, grafici e analisi senza ulteriori operazioni preliminari. Sono tuttavia disponibili alcune opzioni aggiuntive per la preparazione dei dati che potrebbero rivelarsi utili, ad esempio:

- Assegnazione delle proprietà delle variabili che descrivono i dati e determinazione delle modalità di utilizzo di determinati valori;
- Identificazione dei casi che potrebbero contenere informazioni duplicate ed esclusione di tali casi dalle analisi o eliminazione degli stessi dal file di dati;
- Creazione di nuove variabili con categorie distinte che rappresentano gli intervalli di valori delle variabili con un numero elevato di possibili valori.

### 9.1.2 Trasformazioni di dati

Una corretta organizzazione ed elaborazione del DataSet fornisce dati perfettamente adatti al tipo di analisi che si desidera eseguire ed informazioni sulle relazioni tra le variabili. Le analisi preliminari possono essere utili per rivelare la presenza di schemi di codifica difficili o di errori di codifica, oppure può essere necessario eseguire trasformazioni di dati per identificare la relazione effettiva tra le variabili. Il software permette dunque di eseguire trasformazioni di dati rappresentate da attività semplici, ad esempio la compressione delle categorie per l'analisi, o da attività più complesse, come la creazione di nuove variabili in base ad equazioni complesse e istruzioni condizionali.

Utilizzando la finestra di dialogo *Calcola*, è possibile calcolare i valori di una variabile basata su trasformazioni numeriche di altre variabili.

- È possibile calcolare i valori di variabili numeriche o variabili stringa (alfanumeriche);
- È possibile creare nuove variabili oppure sostituire i valori di variabili esistenti. Per le nuove variabili è inoltre possibile specificare il tipo e l'etichetta;
- È possibile calcolare i valori selettivamente per sottoinsiemi di dati in base a condizioni logiche;
- È possibile utilizzare una grande varietà di funzioni predefinite, ad esempi:
  - Funzioni aritmetiche;
  - Funzioni statistiche;
  - Funzioni stringa;
  - Funzioni di data e di ora;
  - Funzioni di distribuzione;
  - Funzioni per variabili casuali;
  - Funzioni per valori mancanti;
  - Funzioni di punteggio.

### 9.1.3 Serie storiche

Sono disponibili numerose trasformazioni di dati che risultano utili per l'analisi delle serie storiche:

- Generazione di variabili della data per definire la periodicità e distinguere tra periodi storici, di convalida e di previsione;
- Creazione di nuove variabili di serie storica come funzioni di variabili di serie storica esistenti;
- Sostituzione di valori mancanti di sistema e definiti dall'utente con stime basate su diversi metodi.

Per ottenere una serie storica è necessario misurare una variabile (o un insieme di variabili) regolarmente per un periodo di tempo. Le trasformazioni di dati di serie storiche si fondano su una struttura del DataSet in cui ciascun caso (riga) rappresenta una serie di osservazioni eseguite in momenti diversi e con un periodo di tempo uniforme tra i casi.

La finestra di dialogo *Crea serie storica* permette di creare nuove variabili basate su funzioni di variabili numeriche di serie storica esistenti. Questi valori trasformati risultano utili in numerose procedure di analisi delle serie storiche. I nuovi nomi di variabile predefiniti sono rappresentati dai primi sei caratteri della variabile esistente in base alla quale sono stati creati, seguiti da un carattere di sottolineatura e da un numero sequenziale. Le nuove variabili mantengono le etichette dei valori definite per le variabili esistenti. Per la creazione di variabili di serie storica sono disponibili diverse funzioni, ad esempio differenze, medie mobili, mediane mobili, ritardo e anticipo.

Il software permette di utilizzare diverse funzioni per la trasformazione delle serie storiche, che sono:

- *Differenza*. Differenza non stagionale tra valori successivi della serie. L'ordine è rappresentato dal numero di valori precedenti utilizzati per calcolare la differenza. Poiché per ciascun ordine di differenza va perduta un'osservazione, i valori mancanti di sistema compaiono all'inizio della serie.
- *Differenza stagionale*. Differenza tra i valori della serie basata su un intervallo di tempo costante. L'intervallo viene definito in base alla periodicità corrente. Per calcolare le differenze stagionali, è innanzitutto necessario definire variabili della data (menu Dati, Definisci dati) che includano un componente periodico (ad esempio i mesi dell'anno). L'ordine è rappresentato dal numero di periodi stagionali utilizzati per calcolare la differenza. Il numero di casi con il valore mancante di sistema all'inizio della serie equivale alla periodicità moltiplicata per l'ordine.
- *Media mobile centrata*. Media di un intervallo di valori di serie che circondano e includono il valore corrente. L'intervallo è rappresentato dal numero di valori di serie utilizzati per calcolare la media. Se l'intervallo è pari, la media mobile verrà calcolata in base alla media di



ciascuna coppia di medie non centrate. Il numero di casi con il valore mancante di sistema all'inizio e alla fine della serie per un intervallo di  $n$  equivale a  $n/2$  per gli intervalli di valori pari e  $(n-1)/2$  per gli intervalli di valori dispari.

- *Media mobile sui precedenti.* Media dell'intervallo di valori di serie che precedono il valore corrente. L'intervallo è rappresentato dal numero di valori di serie precedenti utilizzati per calcolare la media. Il numero di casi con il valore mancante di sistema all'inizio della serie equivale al valore dell'intervallo.
- *Mediane mobili.* Mediane di intervalli di valori di serie che circondano e includono il valore corrente. L'intervallo è rappresentato dal numero di valori di serie utilizzati per calcolare la mediana. Se l'intervallo è pari, la mediana verrà calcolata in base alla media di ciascuna coppia di mediane non centrate. Il numero di casi con il valore mancante di sistema all'inizio e alla fine della serie per un intervallo di  $n$  equivale a  $n/2$  per gli intervalli di valori pari e  $(n-1)/2$  per gli intervalli di valori dispari.
- *Somma cumulata.* La somma cumulata dei valori di serie che precedono e includono il valore corrente.
- *Ritardo.* Valore di un caso precedente, calcolato in base all'ordine di ritardo specificato. L'ordine è rappresentato dal numero di casi che precedono il valore corrente da cui si ottiene tale valore. Il numero di casi con il valore mancante di sistema all'inizio della serie equivale al valore dell'ordine.
- *Anticipo.* Valore di un caso successivo, calcolato in base all'ordine di anticipo specificato. L'ordine è rappresentato dal numero di casi che seguono il caso corrente da cui si ottiene tale valore. Il numero di casi con il valore mancante di sistema alla fine della serie equivale al valore dell'ordine.
- *Livellamento.* Nuovi valori di serie calcolati in base a un algoritmo di livellamento. Il livellatore inizia con la mediana mobile 4, centrata in base alla mediana mobile 2. I valori vengono quindi livellati nuovamente applicando la mediana mobile 5, la mediana mobile 3 e le medie mobili ponderate. I residui vengono calcolati sottraendo le serie livellate dalla serie originale. L'intero processo viene quindi ripetuto sui residui calcolati. Infine, vengono calcolati i residui livellati sottraendo i valori livellati ottenuti nella prima fase del processo. Questo procedimento talvolta viene definito livellamento T4253H.

### 9.1.4 L'output

Quando si esegue una funzione del software, i risultati vengono visualizzati in una finestra denominata *Viewer*. In questa finestra è possibile spostarsi rapidamente verso qualsiasi parte dell'output che si desidera visualizzare. È inoltre possibile gestire l'output e creare un documento che contenga esattamente l'output desiderato.

Il comando *Export Output* salva l'output del *Viewer* nei formati HTML, testo, Word/RTF, Excel, PowerPoint (richiede PowerPoint 97 o versione successiva) e PDF. Anche i grafici possono essere esportati in una vasta gamma di formati.

### 9.1.5 I grafici

Per creare grafici ad alta risoluzione è possibile utilizzare le procedure disponibili nel menu *Grafici* e numerose procedure del menu *Analizza*. In questo capitolo viene fornita una panoramica dello strumento *Grafico*.

Il *Generatore di grafici* permette di creare grafici da modelli di grafici predefiniti o da singole parti (ad esempio assi e barre). Per creare un grafico, è sufficiente trascinare i modelli di grafici o le singole parti nel disegno rappresentato dall'area di grandi dimensioni a destra dell'elenco Variabili nella finestra di dialogo Generatore di grafici. L'area di disegno visualizza un'anteprima del grafico mentre è in corso la creazione. Sebbene l'anteprima utilizzi le etichette di variabile definite e i livelli di misurazione, non visualizza i dati effettivi. Utilizza invece dati generati casualmente per fornire una bozza approssimativa dell'aspetto del grafico.

Sono disponibili molti modelli di grafici che semplificano la scelta da parte dell'utente.

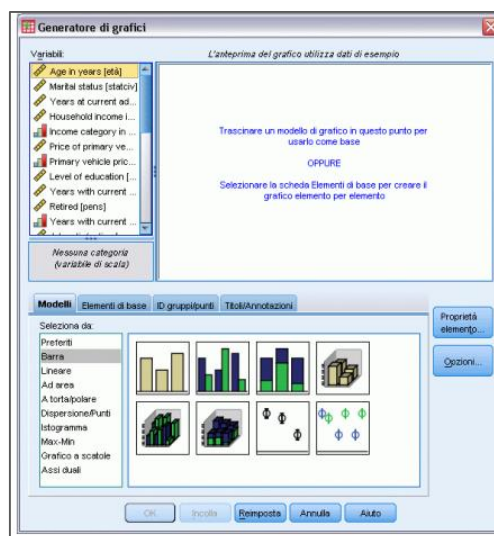


Figura 9.1- Finestra generatore Grafici - IBM® SPSS® Statistics.

### 9.1.6 Le statistiche descrittive dei dati

Per ottenere statistiche descrittive per le variabili selezionate si procede dal menu di scelta rapida selezionare *Statistiche descrittive*. Per impostazione predefinita, le tabelle di frequenza (tabelle dei conteggi) sono visualizzate per tutte le variabili con un massimo di 24 valori univoci. Vengono quindi fornite le seguenti informazioni:

- valore minimo;
- valore massimo;
- media
- mediana
- deviazione standard;
- frequenza di un determinato dato.

È inoltre possibile ottenere grafici a barre per le variabili nominali e ordinali, istogrammi per variabili continue (scala) e modificare il valore cut-off che determina quando visualizzare le tabelle di frequenza.

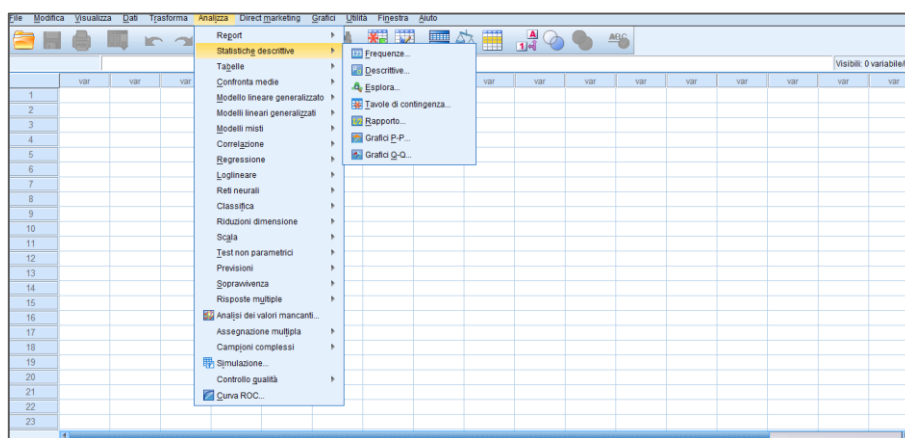


Figura 9.2 - Finestra generatore Analizza - IBM® SPSS® Statistics.

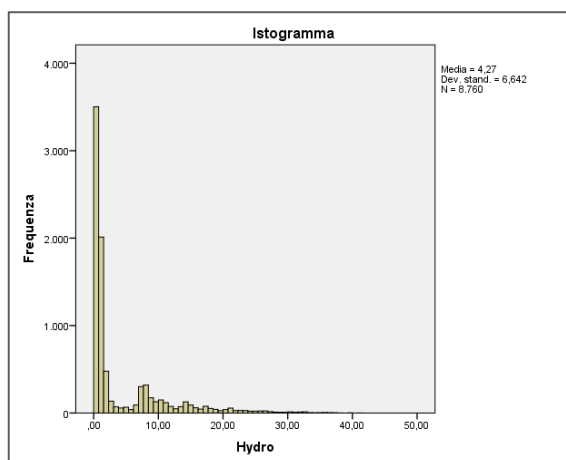


Figura 9.3 - Istogramma tipo per lo studio di frequenza - IBM® SPSS® Statistics.

## **9.2 GMDH Shell (Forecasting Software for Business)**

Nella teoria moderna di modelli predittivi è ben noto che il modello dovrebbe fornire un compromesso tra semplicità e precisione. GMDH Shell raggiunge questo obiettivo utilizzando algoritmi di apprendimento dei modelli.

L'idea di tutti gli algoritmi tipo GMDH è quello di applicare un generatore di modelli che gradualmente ne aumentano la complessità ed, infine, selezionare un insieme di modelli che mostrano la massima precisione nella previsioni future.

Dal punto di vista computazionale, GMDH è abbastanza semplice ed allo stesso tempo è ben coordinato con lo stato dell'arte delle tecniche di data mining. La modalità di validazione rappresenta il cuore centrale per raggiungere una complessità ottimale del modello.

Sono due i tipi di algoritmi di apprendimento:

- Combinatoria GMDH;
- GMDH Neural Network (tipo reti neurali).

Questi due tipi di algoritmo costituiscono gli algoritmi fondamentali del software GMDH Shell, ed hanno il compito di generare il modello migliore. La complessità del modello è infine decisa tramite il metodo dei minimi quadrati.

In generale l'algoritmo fondamentale genera modelli dal più semplice a quello più complesso fino a che la precisione del test di convalida aumenta. Il DataSet è diviso in due sottogruppi, il primo per l'implementazione del modello mentre il secondo per la convalida dello stesso. L'aumento della complessità aumenta gradualmente il numero di parametri interni al modello. Il processo si arresta sotto una soglia minima di diminuzione dell'errore. Il tipo di algoritmi utilizzati e la loro implementazione è stato ben spiegato nei Paragrafi 6.8 e 7.10, pertanto nel proseguo ci concentreremo sul software e le modalità di utilizzo.

### **9.2.1 Il DataSet**

GMDH Shell ha due modi di lettura dei dati:

- CSV / XLS collegamento / XLSX;
- ODBC connessione / OLE DB.

GMDH Shell lavora con serie storiche, cioè sequenze di valori di una variabile a un certo intervallo di tempo. È preferibile avere dati per più di due anni, in caso contrario risultano non rilevanti i fattori di stagionalità poiché non sono in grado di essere valutata. Viceversa, con un DataSet con più di sette anni, il software prende molte più informazioni ed è in grado di valutare la stagionalità ed il rumore

del dato. Relativamente ai valori mancanti, come è noto ognuno dei metodi di previsione statistica richiede un flusso ininterrotto di valori storici; sfortunatamente i valori mancanti sono comuni nei DataSet sperimentali. È opportuno quindi applicare un approccio ben formulato per i dati mancanti.

GMDH Shell ha già predefinito le etichette per segnalare il valore mancante nel DataSet (ad esempio 'N / A', 'NULL'), è inoltre possibile digitare un'etichetta personalizzata.

Nel trattare i dati mancanti GMDH Shell utilizza diversi approcci: sostituirli con media o il valore più utilizzato, o interpolare. È quindi possibile scegliere uno degli approcci.

### 9.2.2 Le analisi sul DataSet

GMDH Shell fornisce i seguenti strumenti per le analisi grafiche dei dati:

*Plot delle serie.* Fornisce una visione dell'andamento della serie storica rilevando periodicità negli andamenti o anomalie di uno o più dati.

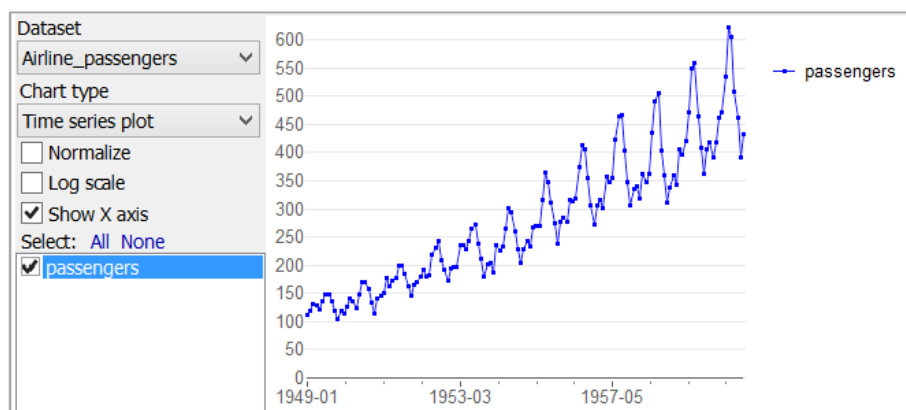


Figura 9.4 - Esempio di plot di serie - GMDH Shell.

*Grafico delle frequenze.* Indica il numero di volte in cui un determinato dato, appartenente alla serie storica si verifica.

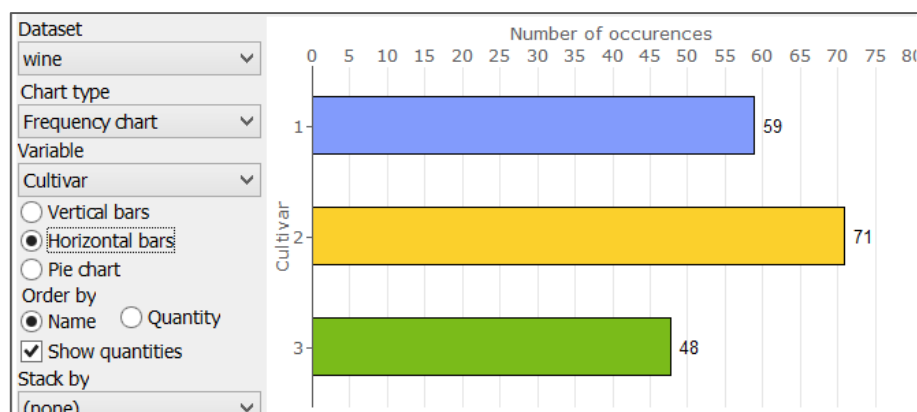


Figura 9.5 - Esempio di grafico delle frequenze - GMDH Shell.

*Scatter. Matrice degli scatter, Scatter in 3D.* Forniscono indicazioni grafiche riguardo l'andamento di una determinata variabile rispetto ad uno o più fattori esterni, facente parte sempre del DataSet.

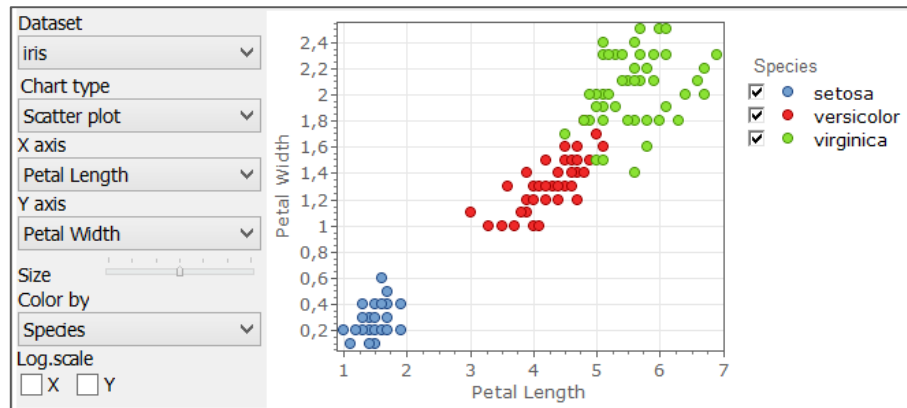


Figura 9.6 Esempio di Scatter - GMDH Shell.

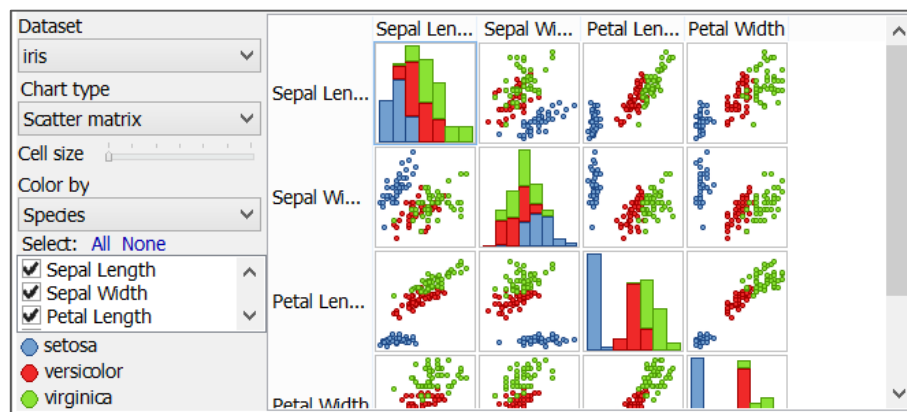


Figura 9.7 - Esempio di matrice Scatter - GMDH Shell.

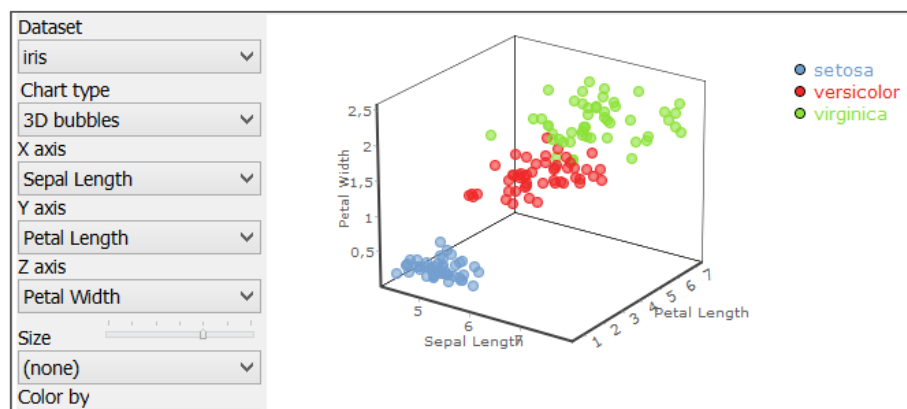


Figura 9.8 - Esempio di Scatter in 3D - GMDH Shell.

*Statistiche del DataSet.* Riporta in opportuna tabella le statistiche descrittive del DataSet.

| Dataset             |                     | Sepal Le... | Sepal Width | Petal Le... | Petal Width | Species |
|---------------------|---------------------|-------------|-------------|-------------|-------------|---------|
| iris                | Numeric values      | 150         | 150         | 150         | 150         | 0       |
| Chart type          | Text values         | 0           | 0           | 0           | 0           | 150     |
| Statistics          | Missing values      | 0           | 0           | 0           | 0           | 0       |
| 150 rows, 5 columns | Unique values       | 35          | 23          | 43          | 22          | 3       |
|                     | Zero values         | 0           | 0           | 0           | 0           | 0       |
|                     | Most frequent       | 5           | 3           | 1,5         | 0,2         | setosa  |
|                     | Min. value          | 4,3         | 2           | 1           | 0,1         |         |
|                     | Max. value          | 7,9         | 4,4         | 6,9         | 2,5         |         |
|                     | Median              | 5,8         | 3           | 4,4         | 1,3         |         |
|                     | Mean value          | 5,843       | 3,054       | 3,759       | 1,199       |         |
|                     | Std. deviation      | 0,8253      | 0,4321      | 1,759       | 0,7606      |         |
|                     | 2 $\sigma$ outliers | 6           | 5           | 0           | 0           |         |

Figura 9.9 - Esempio di tabella delle statistiche - GMDH Shell.

*Istogrammi.* Riporta graficamente la frequenza dei valori del DataSet.

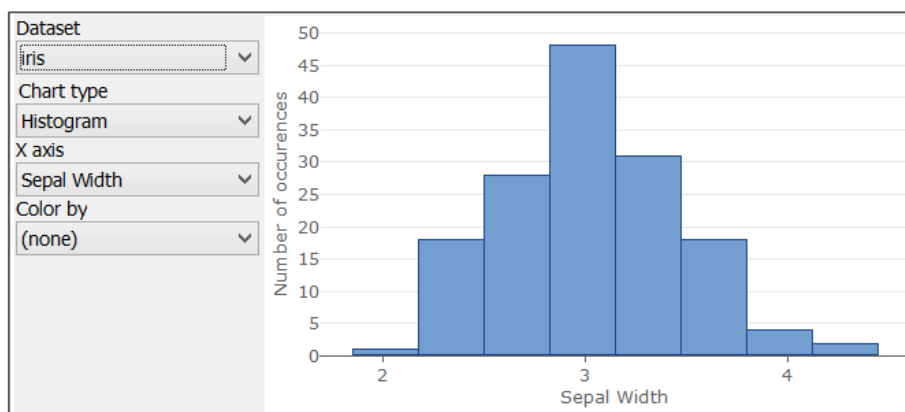


Figura 9.10 - Esempio di istogramma descrittivo - GMDH Shell.

*Autocorrelazione.*

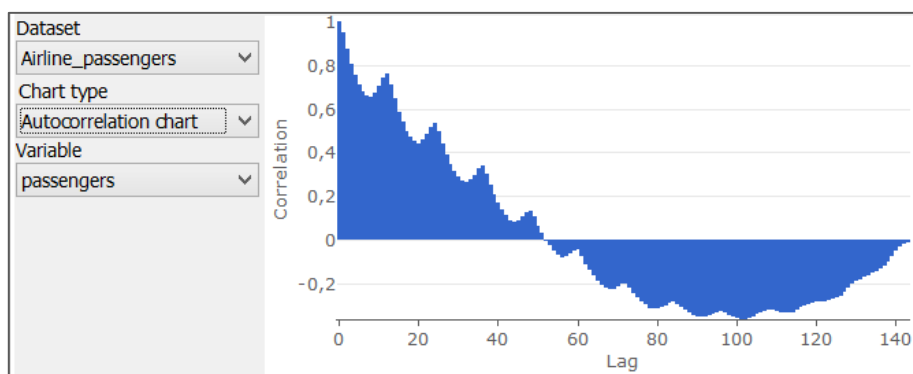


Figura 9.11 - Esempio di grafico delle autocorrelazioni - GMDH Shell.

*Correlazione tra più dati.* Riporta gli indici di correlazioni tra le differenti variabili presenti nel DataSet.

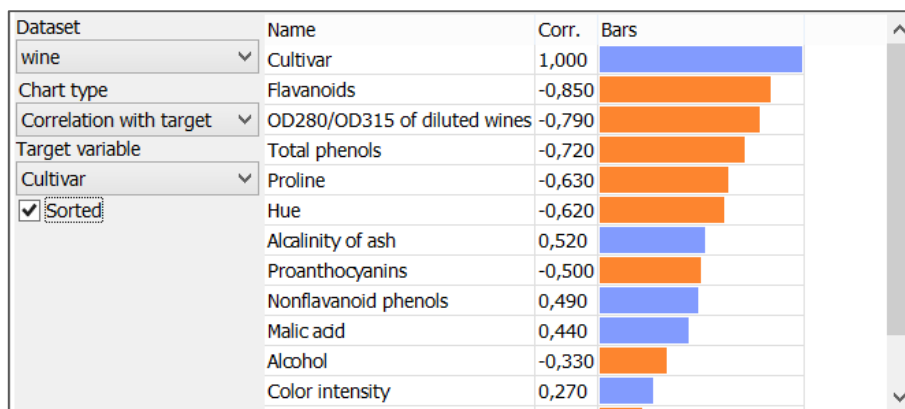


Figura 9.12 - Esempio di tabella delle correlazioni - GMDH Shell.

*Diagramma di correlazione in superficie 3D.*

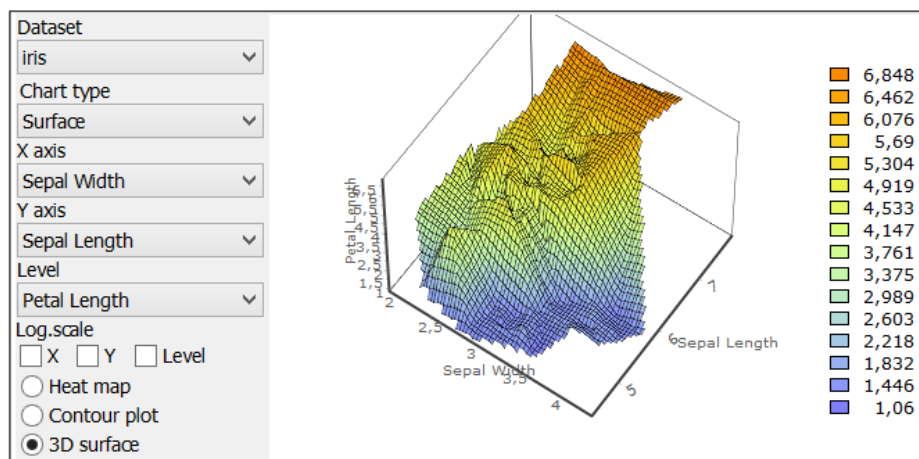


Figura 9.13 - Esempio di diagramma 3D delle correlazioni - GMDH Shell.

*Mappa con gradiente di colore.*

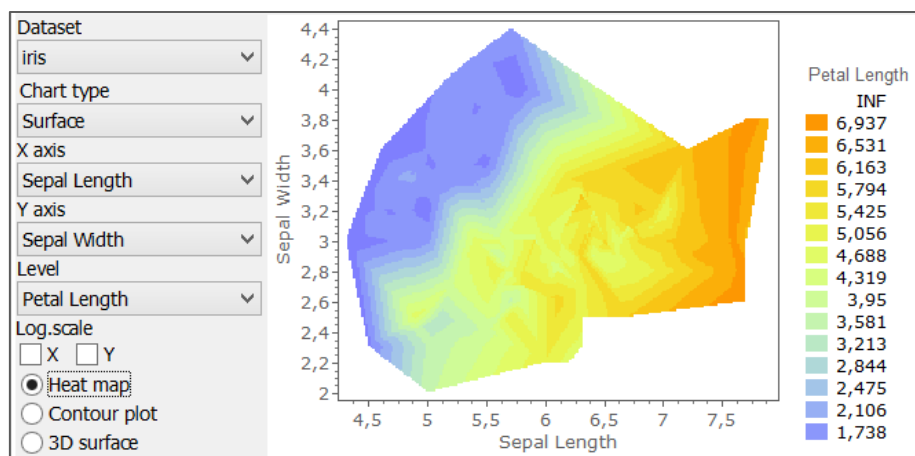


Figura 9.14 - Esempio di mappa a gradienti di colore per le correlazioni - GMDH Shell.



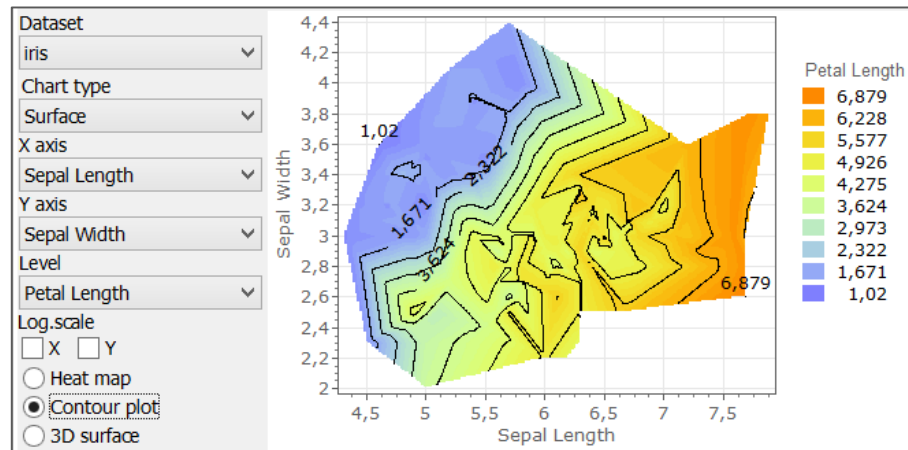


Figura 9.15 - Esempio di mappa a gradienti di colore con curve di valori per le correlazioni - GMDH Shell.

### Grafico OHLC



Figura 9.16 - Esempio di grafico OHLC - GMDH Shell.

### 9.2.3 Pre-processazione del DataSet

GMDH Shell permette la pre-processazione del dato interno al DataSet. Sono disponibili molti tipi di trasformazioni ed il settaggio si genera attraverso il pannello dedicato. È possibile anche impostare a priori delle trasformazioni obbligatorie che il programma genera sul DataSet, queste sono relative alla lettura che deve essere fatta sul dato in formato testo e sul dato mancante.

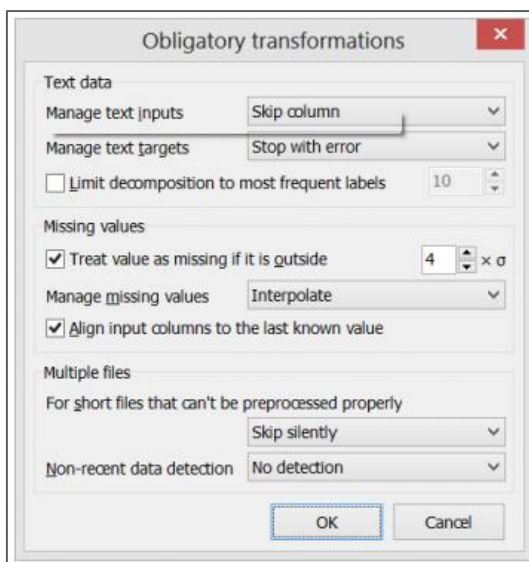


Figura 9.17 - Pannello delle trasformazioni pre-elaborazione - GMDH Shell.

Qualora si vuole limitare il numero di nuove variabili create in seguito a decomposizione, è possibile indicare il limitare di decomposizione alle etichette più frequenti e tutte le altre etichette verranno eliminati. Sono altresì disponibili diverse serie di funzioni di pre-elaborazione per i modelli di serie storiche, la classificazione e modelli di regressione, queste funzioni sono scelte attraverso un apposito pannello e sono:

- *Funzione di previsione semplice*: permette di indicare il numero di osservazioni da prevedere e di valutare l'affidabilità ed accuratezza della previsione con la verifica su dati noti;
- *Funzione previsione di serie temporali*: permette di indicare l'orizzonte o più orizzonti temporali della previsione, valutando l'accuratezza e l'affidabilità con la verifica su dati noti da indicare in apposito settaggio. È possibile anche impostare la ripetizione del calcolo di previsione per lo stesso DataSet;
- *Funzione di regressione e classificazione*: genera due sottoinsiemi del DataSet, al fine di ottimizzare la rete e di verificarne l'affidabilità e l'accuratezza. La grandezza del data set di verifica viene stabilita dall'utente, generalmente il 20% dell'intero DataSet. Può essere infine stabilito se prendere dati appartenenti alla coda del DataSet o uniformi su tutto il DataSet;

Le funzioni di trasformazioni di GMDH Shell sono di seguito riportate.

| Trasformazione    | Esempio Notazione              | Descrizione                                  |
|-------------------|--------------------------------|--|
| Piazza            | x, <code>sqn</code>            | $y = x^2$                                    |
| Radice quadrata   | x, <code>sqrt</code>           | $y = x^{(1/2)}$                              |
| Cubo              | x, <code>il cubo</code>        | $y = x^3$                                    |
| radice cubica     | x, <code>Cubert</code>         | $y = x^{(1/3)}$                              |
| Exp               | x, <code>exp</code>            | $y = \exp(x)$                                |
| Logaritmo         | x, <code>ln</code>             | $y = \ln(x \mid  ), x > 0$                   |
| Seno              | x, <code>il peccato</code>     | $y = \sin(x)$                                |
| Coseno            | x, <code>cos</code>            | $y = \cos(x)$                                |
| Arcotangente      | x, <code>arctang</code>        | $y = \arctang(x)$                            |
| Valore Abs        | x, <code>abs</code>            | $y =  x $                                    |
| Segno             | x, <code>segno</code>          | $y = \text{segno}(x)$                        |
| Pavimento         | x, <code>piano</code>          | $y = [x]$                                    |
| Parte frazionaria | x, <code>frac</code>           | $y = x - [x]$                                |
| Normalizzazione   | x, <code>norma (b1, b2)</code> | b1 è limite inferiore, b2 è limite superiore |

Tabella 9.1 - Funzioni elementari - GMDH Shell.

| Trasformazione      | Esempio Notazione                |
|---------------------|----------------------------------|
| Ritardo             | x @ ab: c                        |
| Tutti i ritardi     | x @ *: una                       |
| Media mobile        | x, <code>SMA (a)</code>          |
| Esponenziale MA     | x, <code>EMA (a)</code>          |
| Derivato            | x, <code>d</code>                |
| Dimensioni finestra | x, <code>finestra (a)</code>     |
| Ponderata nel tempo | x, <code>weighted_by_time</code> |
| Serie di Fourier    | x, <code>Fourier (a)</code>      |

Tabella 9.2 - Funzioni per serie temporali - GMDH Shell.

| Trasformazione         | Esempio Notazione                                     |
|------------------------|---|
| Anno                   | <code>x, anno</code>                                  |
| Frazione Anno          | <code>x, year_frac</code>                             |
| Mese                   | <code>x, il mese,</code><br><code>decomporre</code>   |
| Frazione Mese          | <code>x, month_frac</code>                            |
| Giorno del mese        | <code>x, giorno,</code><br><code>decomporre</code>    |
| Giorno della settimana | <code>x, dayofweek,</code><br><code>decomporre</code> |
| Frazione Giorno        | <code>x, day_frac</code>                              |
| Ora                    | <code>x, ora, decomporre</code>                       |
| Frazione ora           | <code>x, hour_frac</code>                             |
| Minuto                 | <code>x, minuto,</code><br><code>decomporre</code>    |
| Secondo                | <code>x, secondo,</code><br><code>decomporre</code>   |

Tabella 9.3 - Funzioni per set di dati temporali - GMDH Shell.

| Trasformazione                | Esempio Notazione                                 |
|-------------------------------|---|
| Giorno libero                 | <code>x, isdayoff</code>                          |
| Giorni lavorativi per periodo | <code>x,</code><br><code>workdaysperperiod</code> |
| Vacanze decomposto            | <code>x, vacanze</code>                           |
| È vacanza                     | <code>x, isholiday</code>                         |
| Trova non festivo             | <code>x, nonholiday_day</code>                    |

Tabella 9.4 - Funzioni calendario - GMDH Shell.

| Trasformazione                  | Esempio Notazione  |
|---------------------------------|--|
| Ponderata nel tempo             | <code>x,</code><br><code>weighted_by_time</code>                         |
| Classi equilibrate              | <code>x,</code><br><code>balanced_classes</code>                         |
| Manuale polarizzazione 2 classe | <code>x,</code><br><code>weighted_two_class</code><br><code>(0.5)</code> |

Tabella 9.5 - Funzioni per calibrazione - GMDH Shell.

| Trasformazione | Esempio Notazione  |
|----------------|--------------------|
| Orario         | Tempo              |
| Bersaglio      | bersaglio          |
| ID             | id                 |
| Prima colonna  | firstColumn<br>(a) |
| Ultima colonna | lastColumn<br>(a)  |

Tabella 9.6 - Variabili speciali - GMDH Shell.

### 9.2.4 Il Set del risolutore

Il modulo Solver del GMDH Shell permette di settare le impostazioni per la realizzazione dei modelli predittivi per le variabili target. Il modulo si presenta come in Figura 9.18.

The Solver module interface includes the following settings:

- Reorder observations: No
- Validation strategy: k-fold validation
- Number of folds: 2
- Validation criterion: RMSE · balance
- Variables ranking: By error (independe
- Drop variables after rank: 600
- Core algorithm: Neural-type
- Neuron function:  $a + x_i + x_i \cdot x_j + \dots$
- Max. number of layers: 33
- Initial layer width: 2 000
- ☐ Set parallel threads manually: 4

Figura 9.18 - Modulo risolutore - GMDH Shell.

Dove:

- *Reorder observations*: è utilizzato per ottenere caratteristiche statistiche uniformi.
- *Validation strategy*: setta le modalità di convalida del modello. Le opzioni sono quattro ed il dettaglio è riportato in Tabella 9.7. Per ogni opzione si deve indicare il numero o la percentuale di dati che compongono l'insieme di formazione e quello che compone l'insieme di test.

| Opzioni            | Descrizione   |
|--------------------|---|
| Training/testing   | Suddivide il DataSet in due parti, la prima è utilizzata per la formazione del modello, la seconda è utilizzata per il confronto tra i modelli generati al fine di selezione quello che ha le migliori prestazioni. |
| Whole data testing | Divide il set di dati in due sottoinsiemi, la formazione è svolta sul primo mentre il test su tutto il DataSet.   |
| k-fold validation  | Divide il set di dati in n insiemi, ogni insieme è quindi suddiviso in due sottoinsiemi che sono utilizzati per la formazione ed il test.   |
| Leave-one-out CV   | Rappresenta una k-fold cross-validation dove il numero di n insiemi è pari al numero delle osservazioni presenti sul DataSet.   |

Tabella 9.7 - Modalità di convalida del modello - GMDH Shell.

- *Validation criterion*: l'utente tramite il settaggio determina il criterio di selezione del modello, e quindi dell'algoritmo, rispetto a quelli generati. Il tipo di criterio può essere: RMSE, MAE, RMSE balance.
- *Variables ranking*: permette di indicare il numero massimo di parametri indipendenti che devono comporre l'algoritmo del modello.
- *Core algorithm*: l'utente può decidere di utilizzare un algoritmo combinatorio o una Neural Network di tipo GMDH. In base al settaggio si stabilisce la complessità computazionale dell'algoritmo, o il numero di layer nascosti e il numero di neuroni aggiunti ad ogni nuovo layer per Reti-Neurali.

### 9.2.5 Il post-processor

Il Post-processor viene utilizzato per trasformare le previsioni dei modelli dopo il completamento di una simulazione. Non appena i parametri di post-processo vengono modificati, tutte le previsioni e gli altri risultati della simulazione vengono aggiornati in tempo reale.

L'opzione "*Average of top-ranked models*" è utilizzata per ridurre la varianza dei residui. Tuttavia la selezione a posteriori del numero di top-ranked modelli potrebbe portare a un eccesso di montaggio e bassa accuratezza delle previsioni.

Sono presenti altre due opzioni, la prima sostituisce i valori della previsione all'intero più vicino, mentre la seconda sostituisce i valori negativi con valori pari a zero.

### 9.2.6 L'analisi dei risultati

GMDH permette di visualizzare i risultati delle simulazioni sia tramite plot che tramite tabella, come si può vedere dalla Figura 9.19.

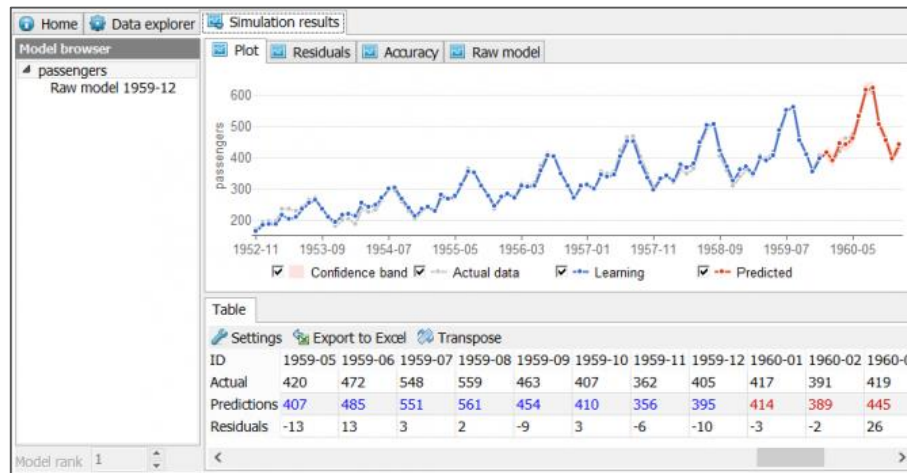


Figura 9.19 - Finestra dei risultati - GMDH Shell.

Nel plot i valori in blu rappresentano i valori del fit, formazione del modello, mentre in rosso i valori previsti. In grigio sono riportati i dati reali mentre l'area in rosso rappresenta la banda di confidenza al 95%. È quindi possibile visualizzare il modello generato tramite l'opzione *Raw model*. Il pannello fornisce informazioni sull'Id del Modello, la sua complessità ed il valore del criterio di convalida.

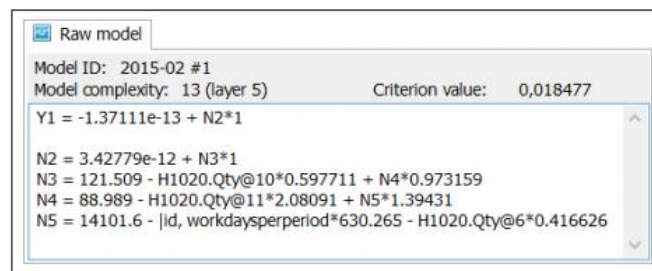


Figura 9.20 - Finestra del modello - GMDH Shell.

Il pannello *Accuracy* fornisce informazioni sul modello per quanto riguarda la sua precisione ed affidabilità. In primo luogo sono distinti i due insiemi: *Model fit* e *Predictions*. Per ognuno è fornito un ben preciso dato. Gli errori sono quindi organizzati per tre matrici differenti: *Assoluto*, *Range Percentage* e *Target percentage*.

Sono forniti i seguenti indicatori:

- Max. negative error;
- Max. positive error;
- Residual sum;
- Standard deviation of residuals;
- Coefficient of determination ( $R^2$ );
- Correlation coefficient .

In fine GMDH Shell fornisce l'importanza delle variabili indipendenti all'interno del modello. L'importanza della variabile è calcolata sostituendo il valore medio della stessa variabile al modello e calcolando il RMSE. Può capitare che il valore sia in percentuale maggiore del 100%, questo avviene quando nel modello la variabile è moltiplicata per un'altra variabile o è elevata ad un esponente.



# CAPITOLO 10 *Caso Studio*

## *Consumi energetici*

### 10.1 Scala regionale

Lo studio degli indici su scala regionale è stato effettuato per la regione Sicilia. Il data base rappresentativo di dati annuali riporta indicatori energetici, ambientali, demografici ed economici della regione. I vettori energetici analizzati si riferiscono ai combustibili solidi, liquidi, gassosi ed all'energia elettrica. Tuttavia, poiché i dati storici per i combustibili sono allineabili esclusivamente per il triennio 2010-2012 non è possibile strutturare con così pochi dati (3 anni) un modello predittivo affidabile.

Pertanto per il caso studio su scala regionale si è svolta un'analisi predittiva dei consumi di energia elettrica per singolo settore (residenziale, servizi non vendibili, agricoltura industria, terziario ed attività produttive). Tale analisi è stata svolta con una base dati di 23 anni (1990-2012) sia con modello su serie storica che con modello Neural Network. In quest'ultimo caso come variabili di input sono stati scelti i valori aggiunti dei singoli settori, il numero della popolazione e la temperatura media annuale.

#### 10.1.1 *Il DataSet*

Tramite i dati provenienti dal Ministero dello Sviluppo Economico [36-37], TERNA Spa [38], GSE Spa [39-42], ISPRA [43], Istat [44] ed Istituto Guglielmo Tagliacarne [45] è stato possibile la costruzione del database regionale che si riporta nelle Tabelle 10.1-10.4

| Anno | T media [°C] | Popolazione | Valore_agg_Agricoltura [€] | Valore_agg_Industria [€] | Valore_agg_Terziario [€] | Valore_agg_Attività_produttive [€] | Co <sub>2</sub> _fatt [tCO <sub>2</sub> /tep] |
|------|--------------|-------------|----------------------------|--------------------------|--------------------------|------------------------------------|---|
| 1990 | 17,31        | 4969977     | 2930587280                 | 8207791571               | 27736556381              | 38874935232                        | 6,88  |
| 1991 | 16,36        | 4966386     | 3121613663                 | 8741249864               | 29858491237              | 41721354764                        | 6,81  |
| 1992 | 16,75        | 4967386     | 3092450697                 | 8946761050               | 31540542280              | 43579754027                        | 6,74  |
| 1993 | 16,86        | 4968386     | 3063287731                 | 9152272236               | 33222593324              | 45438153290                        | 6,67  |
| 1994 | 18,06        | 4969385     | 3034124765                 | 9357783421               | 34904644367              | 47296552553                        | 6,60  |
| 1995 | 16,95        | 4970385     | 3004961799                 | 9563294607               | 36586695410              | 49154951816                        | 6,53  |
| 1996 | 16,88        | 4971385     | 2975798833                 | 9768805793               | 38268746454              | 51013351079                        | 6,41  |
| 1997 | 16,94        | 4972385     | 2946635867                 | 9974316978               | 39950797497              | 52871750343                        | 6,28  |
| 1998 | 16,94        | 4973385     | 2863723760                 | 10179828164              | 41632848541              | 54676400465                        | 6,16  |
| 1999 | 18,27        | 4974384     | 2780811654                 | 10385339350              | 43314899584              | 56481050587                        | 6,03  |
| 2000 | 17,95        | 4975384     | 2697899547                 | 10590850535              | 44996950628              | 58285700710                        | 5,91  |
| 2001 | 17,93        | 4968991     | 2614987440                 | 10796361721              | 46679001671              | 60090350832                        | 5,77  |

|      |       |         |             |              |              |              |      |
|------|-------|---------|-------------|--------------|--------------|--------------|------|
| 2002 | 17,56 | 4979781 | 2618232782  | 10894946705  | 48448217093  | 61961396580  | 5,94 |
| 2003 | 17,37 | 4990571 | 2621478123  | 10993531690  | 50217432515  | 63832442328  | 5,86 |
| 2004 | 16,65 | 5001361 | 2624723464  | 11092116674  | 51986647938  | 65703488076  | 5,59 |
| 2005 | 16,18 | 5012152 | 2627968806  | 11190701658  | 53755863360  | 67574533824  | 5,60 |
| 2006 | 16,98 | 5022942 | 2631214147  | 11289286643  | 55525078782  | 69445579572  | 5,51 |
| 2007 | 17,02 | 5033732 | 2634459488  | 11387871627  | 57294294204  | 71316625319  | 5,34 |
| 2008 | 17,14 | 5044522 | 2645917461  | 11486456612  | 59063509626  | 73195883699  | 5,21 |
| 2009 | 17,03 | 5040397 | 2657375434  | 11585041596  | 60832725048  | 75075142078  | 4,80 |
| 2010 | 16,57 | 5042992 | 2719898432  | 11232020798  | 61622362524  | 75574281754  | 4,60 |
| 2011 | 16,73 | 5050625 | 2839000000* | 10879000000* | 62412000000* | 76130000000* | 4,57 |
| 2012 | 17,12 | 4999932 | 3055000000* | 10122000000* | 60804000000* | 73981000000* | 4,38 |

Tabella 10.1 – Valori climatici, economici ed ambientali DataSet Sicilia.

(\* dato disaggregato a partire dal dato totale di Valore aggiunto Sicilia 2011 e 2012)

| Anno | Residenziale<br>[ktep] | Servizi_non_vendibili<br>[ktep] | Agricoltura<br>[ktep] | Industria<br>[ktep] | Terziario<br>[ktep] | Attività_produttive<br>[ktep] |
|------|------------------------|---------------------------------|-----------------------|---------------------|---------------------|-------------------------------|
| 1990 | 418,79                 | 139,35                          | 34,27                 | 508,14              | 127,72              | 670,13                        |
| 1991 | 449,66                 | 151,82                          | 32,28                 | 509,04              | 135,17              | 676,49                        |
| 1992 | 466,55                 | 160,61                          | 31,71                 | 528,51              | 140,27              | 700,50                        |
| 1993 | 468,07                 | 164,46                          | 39,50                 | 510,78              | 144,87              | 695,15                        |
| 1994 | 462,34                 | 167,66                          | 36,05                 | 502,10              | 150,78              | 688,93                        |
| 1995 | 463,96                 | 144,40                          | 32,12                 | 511,42              | 180,60              | 724,14                        |
| 1996 | 458,86                 | 145,34                          | 28,63                 | 494,11              | 181,47              | 704,21                        |
| 1997 | 463,41                 | 149,16                          | 28,68                 | 505,53              | 189,87              | 724,09                        |
| 1998 | 458,09                 | 155,32                          | 31,70                 | 512,18              | 199,10              | 742,98                        |
| 1999 | 473,01                 | 164,05                          | 36,98                 | 557,81              | 213,05              | 807,84                        |
| 2000 | 464,56                 | 170,38                          | 35,56                 | 603,01              | 222,19              | 860,77                        |
| 2001 | 464,42                 | 173,74                          | 38,61                 | 613,55              | 233,83              | 885,99                        |
| 2002 | 473,97                 | 177,72                          | 38,94                 | 629,74              | 245,96              | 914,64                        |
| 2003 | 487,58                 | 186,26                          | 34,57                 | 614,26              | 260,90              | 909,73                        |
| 2004 | 502,77                 | 184,07                          | 34,09                 | 595,67              | 271,81              | 901,57                        |
| 2005 | 509,73                 | 197,56                          | 35,02                 | 585,30              | 287,84              | 908,16                        |
| 2006 | 510,34                 | 202,50                          | 36,93                 | 587,33              | 311,59              | 935,84                        |
| 2007 | 508,11                 | 196,95                          | 37,47                 | 575,52              | 322,91              | 935,90                        |
| 2008 | 504,48                 | 202,16                          | 38,35                 | 560,02              | 334,73              | 933,10                        |
| 2009 | 505,23                 | 197,86                          | 34,96                 | 519,21              | 339,85              | 894,02                        |
| 2010 | 502,95                 | 198,38                          | 34,82                 | 554,35              | 350,97              | 940,14                        |
| 2011 | 516,23                 | 193,05                          | 34,62                 | 563,37              | 358,45              | 956,43                        |
| 2012 | 519,20                 | 191,90                          | 36,03                 | 520,78              | 373,64              | 930,46                        |

Tabella 10.2 – Consumi energetici DataSet Sicilia.

| Anno | Residenziale<br>[ktep/p.ne] | Servizi_non_vendibili<br>[ktep/Mln.€] | Agricoltura<br>[ktep/Mln.€] | Industria<br>[ktep/Mln.€] | Terziario<br>[ktep/Mln.€] | Attività_produttive<br>[ktep/Mln.€] |
|------|-----------------------------|---------------------------------------|-----------------------------|---------------------------|---------------------------|-------------------------------------|
| 1990 | 0,08                        | 0,03                                  | 11,69                       | 61,91                     | 4,60                      | 78,21                               |
| 1991 | 0,09                        | 0,03                                  | 10,34                       | 58,23                     | 4,53                      | 73,10                               |
| 1992 | 0,09                        | 0,03                                  | 10,25                       | 59,07                     | 4,45                      | 73,77                               |
| 1993 | 0,09                        | 0,03                                  | 12,89                       | 55,81                     | 4,36                      | 73,06                               |
| 1994 | 0,09                        | 0,03                                  | 11,88                       | 53,66                     | 4,32                      | 69,86                               |
| 1995 | 0,09                        | 0,03                                  | 10,69                       | 53,48                     | 4,94                      | 69,10                               |
| 1996 | 0,09                        | 0,03                                  | 9,62                        | 50,58                     | 4,74                      | 64,94                               |
| 1997 | 0,09                        | 0,03                                  | 9,73                        | 50,68                     | 4,75                      | 65,17                               |
| 1998 | 0,09                        | 0,03                                  | 11,07                       | 50,31                     | 4,78                      | 66,16                               |
| 1999 | 0,10                        | 0,03                                  | 13,30                       | 53,71                     | 4,92                      | 71,93                               |
| 2000 | 0,09                        | 0,03                                  | 13,18                       | 56,94                     | 4,94                      | 75,06                               |
| 2001 | 0,09                        | 0,03                                  | 14,76                       | 56,83                     | 5,01                      | 76,60                               |
| 2002 | 0,10                        | 0,04                                  | 14,87                       | 57,80                     | 5,08                      | 77,75                               |

|      |      |      |       |       |      |       |
|------|------|------|-------|-------|------|-------|
| 2003 | 0,10 | 0,04 | 13,19 | 55,88 | 5,20 | 74,26 |
| 2004 | 0,10 | 0,04 | 12,99 | 53,70 | 5,23 | 71,92 |
| 2005 | 0,10 | 0,04 | 13,33 | 52,30 | 5,35 | 70,98 |
| 2006 | 0,10 | 0,04 | 14,03 | 52,03 | 5,61 | 71,67 |
| 2007 | 0,10 | 0,04 | 14,22 | 50,54 | 5,64 | 70,40 |
| 2008 | 0,10 | 0,04 | 14,49 | 48,76 | 5,67 | 68,92 |
| 2009 | 0,10 | 0,04 | 13,16 | 44,82 | 5,59 | 63,56 |
| 2010 | 0,10 | 0,04 | 12,80 | 49,35 | 5,70 | 67,85 |
| 2011 | 0,10 | 0,04 | 12,19 | 51,78 | 5,74 | 69,72 |
| 2012 | 0,10 | 0,04 | 11,80 | 51,45 | 6,15 | 69,39 |

Tabella 10.3 – Indicatori energetici DataSet Sicilia.

| Anno | Residenziale<br>[tCO <sub>2</sub> /p.ne] | Servizi_non_vendibili<br>[tCO <sub>2</sub> /Mln.€] | Agricoltura<br>[tCO <sub>2</sub> /Mln.€] | Industria<br>[tCO <sub>2</sub> /Mln.€] | Terziario<br>[tCO <sub>2</sub> /Mln.€] | Attività_produttive<br>[tCO <sub>2</sub> /Mln.€] |
|------|--|--|--|--|--|--|
| 1990 | 0,58                                     | 0,19   | 80,50                                    | 426,17                                 | 31,70                                  | 538,36   |
| 1991 | 0,62                                     | 0,21   | 70,45                                    | 396,81                                 | 30,85                                  | 498,11   |
| 1992 | 0,63                                     | 0,22   | 69,15                                    | 398,40                                 | 29,99                                  | 497,55   |
| 1993 | 0,63                                     | 0,22   | 86,06                                    | 372,49                                 | 29,10                                  | 487,66   |
| 1994 | 0,61                                     | 0,22   | 78,48                                    | 354,38                                 | 28,53                                  | 461,39   |
| 1995 | 0,61                                     | 0,19   | 69,85                                    | 349,47                                 | 32,26                                  | 451,58   |
| 1996 | 0,59                                     | 0,19   | 61,66                                    | 324,19                                 | 30,39                                  | 416,24   |
| 1997 | 0,59                                     | 0,19   | 61,16                                    | 318,48                                 | 29,86                                  | 409,51   |
| 1998 | 0,57                                     | 0,19   | 68,17                                    | 309,84                                 | 29,45                                  | 407,45   |
| 1999 | 0,57                                     | 0,20   | 80,22                                    | 324,02                                 | 29,67                                  | 433,91   |
| 2000 | 0,55                                     | 0,20   | 77,86                                    | 336,33                                 | 29,17                                  | 443,36   |
| 2001 | 0,54                                     | 0,20   | 85,15                                    | 327,76                                 | 28,89                                  | 441,80   |
| 2002 | 0,57                                     | 0,21   | 88,37                                    | 343,45                                 | 30,17                                  | 461,99   |
| 2003 | 0,57                                     | 0,22   | 77,29                                    | 327,45                                 | 30,45                                  | 435,19   |
| 2004 | 0,56                                     | 0,21   | 72,64                                    | 300,36                                 | 29,24                                  | 402,24   |
| 2005 | 0,57                                     | 0,22   | 74,69                                    | 293,14                                 | 30,01                                  | 397,83   |
| 2006 | 0,56                                     | 0,22   | 77,35                                    | 286,74                                 | 30,93                                  | 395,03   |
| 2007 | 0,54                                     | 0,21   | 75,91                                    | 269,73                                 | 30,08                                  | 375,72   |
| 2008 | 0,52                                     | 0,21   | 75,50                                    | 253,98                                 | 29,52                                  | 359,00   |
| 2009 | 0,48                                     | 0,19   | 63,18                                    | 215,23                                 | 26,83                                  | 305,23   |
| 2010 | 0,46                                     | 0,18   | 58,95                                    | 227,26                                 | 26,23                                  | 312,44   |
| 2011 | 0,47                                     | 0,17   | 55,73                                    | 236,71                                 | 26,25                                  | 318,69   |
| 2012 | 0,45                                     | 0,17   | 51,66                                    | 225,35                                 | 26,92                                  | 303,93   |

Tabella 10.4 – Indicatori ambientali DataSet Sicilia.

### 10.1.2 Analisi del DataSet

Le statistiche descrittive dei dati sono riportate nelle Tabelle 10.5-10.8

|                     | T media<br>[°C] | Popolazione | Valore_agg_<br>Agricoltura | Valore_agg_<br>Industria | Valore_agg_<br>Terziario | Valore_agg_<br>Attività_produttive | Co <sub>2</sub> _fatt<br>[tCO <sub>2</sub> /tep] |
|---------------------|-----------------|-------------|----------------------------|--------------------------|--------------------------|------------------------------------|--|
| Numero vaolri       | 23,00           | 23,00       | 23,00                      | 23,00                    | 23,00                    | 23,00                              | 23,00  |
| Valori di testo     | 0,00            | 0,00        | 0,00                       | 0,00                     | 0,00                     | 0,00                               | 0,00   |
| Valori mancanti     | 0,00            | 0,00        | 0,00                       | 0,00                     | 0,00                     | 0,00                               | 0,00   |
| Vaoliri unici       | 22,00           | 23,00       | 23,00                      | 23,00                    | 23,00                    | 23,00                              | 23,00  |
| Valori nulli        | 0,00            | 0,00        | 0,00                       | 0,00                     | 0,00                     | 0,00                               | 0,00   |
| Valore minimo       | 16,18           | 4966386     | 2614987440                 | 8207791570               | 27736556381              | 38874935231                        | 4,38   |
| Valore massimo      | 18,27           | 5050625     | 3121613662                 | 11585041596              | 62412000000              | 76130000000                        | 6,88   |
| Media               | 16,98           | 4975384     | 2780811653                 | 10590850535              | 46679001671              | 60090350832                        | 5,91   |
| Valore medio        | 17,11           | 4994209     | 2817484833                 | 10339896925              | 46550212976              | 59707594735                        | 5,84   |
| Deviazione standard | 0,53            | 29479       | 182349705                  | 951038257                | 10947712835              | 11683573572                        | 0,74   |

Tabella 10.5 – Statistiche descrittive valori climatici, economici ed ambientali DataSet Sicilia.

|                     | Residenziale | Servizi_non_vendibili | Agricoltura | Industria | Terziario | Attività_produttive |
|---------------------|--------------|-----------------------|-------------|-----------|-----------|---------------------|
| Numero vaolri       | 23,00        | 23,00                 | 23,00       | 23,00     | 23,00     | 23,00               |
| Valori di testo     | 0,00         | 0,00                  | 0,00        | 0,00      | 0,00      | 0,00                |
| Valori mancanti     | 0,00         | 0,00                  | 0,00        | 0,00      | 0,00      | 0,00                |
| Vaoliri unici       | 23,00        | 23,00                 | 23,00       | 23,00     | 23,00     | 23,00               |
| Valori nulli        | 0,00         | 0,00                  | 0,00        | 0,00      | 0,00      | 0,00                |
| Valore minimo       | 418,79       | 139,35                | 28,63       | 494,11    | 127,72    | 670,13              |
| Valore massimo      | 519,20       | 202,50                | 39,50       | 629,74    | 373,64    | 956,43              |
| Media               | 473,01       | 173,74                | 34,96       | 554,35    | 233,83    | 885,99              |
| Valore medio        | 480,54       | 174,55                | 34,86       | 550,51    | 242,50    | 827,88              |
| Deviazione standard | 25,63        | 20,40                 | 2,92        | 42,02     | 78,45     | 105,17              |

Tabella 10.6 – Statistiche descrittive consumi energetici DataSet Sicilia.

| Anno                | Residenziale | Servizi_non_vendibili | Agricoltura | Industria | Terziario | Attività_produttive |
|---------------------|--------------|-----------------------|-------------|-----------|-----------|---------------------|
| Numero vaolri       | 23,00        | 23,00                 | 23,00       | 23,00     | 23,00     | 23,00               |
| Valori di testo     | 0,00         | 0,00                  | 0,00        | 0,00      | 0,00      | 0,00                |
| Valori mancanti     | 0,00         | 0,00                  | 0,00        | 0,00      | 0,00      | 0,00                |
| Vaoliri unici       | 23,00        | 23,00                 | 23,00       | 23,00     | 23,00     | 23,00               |
| Valori nulli        | 0,00         | 0,00                  | 0,00        | 0,00      | 0,00      | 0,00                |
| Valore minimo       | 0,08         | 0,03                  | 9,62        | 44,82     | 4,32      | 63,56               |
| Valore massimo      | 0,10         | 0,04                  | 14,87       | 61,91     | 6,15      | 78,21               |
| Media               | 0,10         | 0,03                  | 12,89       | 53,48     | 5,01      | 70,98               |
| Valore medio        | 0,10         | 0,03                  | 12,46       | 53,46     | 5,10      | 71,02               |
| Deviazione standard | 0,00         | 0,00                  | 1,56        | 3,83      | 0,50      | 3,89                |

Tabella 10.7 – Statistiche indicatori energetici DataSet Sicilia.

| Anno                | Residenziale | Servizi_non_vendibili | Agricoltura | Industria | Terziario | Attività_produttive |
|---------------------|--------------|-----------------------|-------------|-----------|-----------|---------------------|
| Numero vaolri       | 23,00        | 23,00                 | 23,00       | 23,00     | 23,00     | 23,00               |
| Valori di testo     | 0,00         | 0,00                  | 0,00        | 0,00      | 0,00      | 0,00                |
| Valori mancanti     | 0,00         | 0,00                  | 0,00        | 0,00      | 0,00      | 0,00                |
| Vaoliri unici       | 23,00        | 23,00                 | 23,00       | 23,00     | 23,00     | 23,00               |
| Valori nulli        | 0,00         | 0,00                  | 0,00        | 0,00      | 0,00      | 0,00                |
| Valore minimo       | 0,45         | 0,17                  | 51,66       | 215,23    | 26,23     | 303,93              |
| Valore massimo      | 0,63         | 0,22                  | 88,37       | 426,17    | 32,26     | 538,36              |
| Media               | 0,57         | 0,20                  | 74,69       | 324,02    | 29,67     | 416,24              |
| Valore medio        | 0,56         | 0,20                  | 72,17       | 313,81    | 29,41     | 415,40              |
| Deviazione standard | 0,05         | 0,02                  | 9,62        | 56,29     | 1,56      | 63,25               |

Tabella 10.8 – Statistiche indicatori ambientali DataSet Sicilia.

Attraverso la matrice degli scatter è possibile verificare le correlazioni tra le differenti variabili del DataSet, (Figura 10.1).

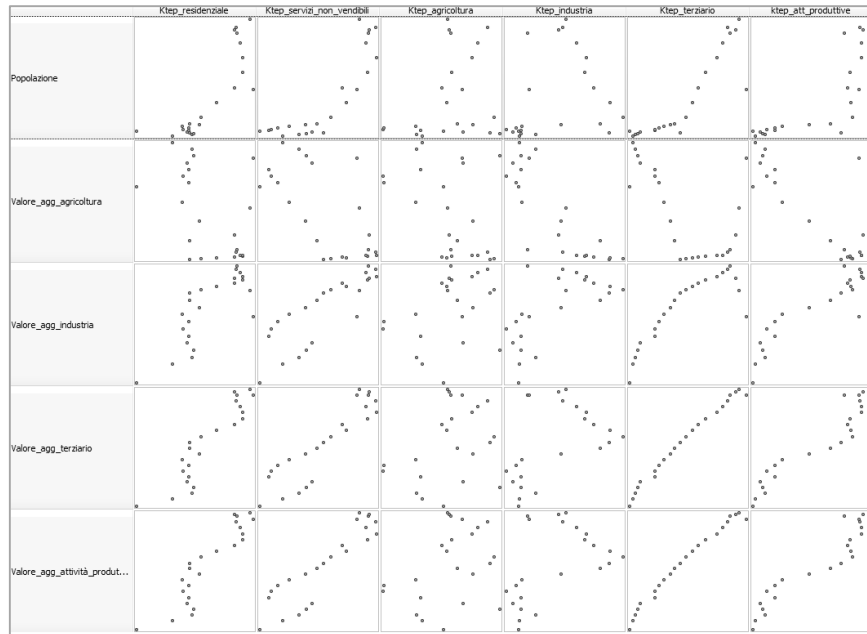


Figura 10.1 – Matrice degli scatter DataSet Sicilia.

È stato quindi verificato il valore dell'indice di correlazione per il singolo target (Figure 10.2 - 10.8). Il valore di indice positivo indica che il valore targhet cresce con il crescere del valore del fattore correlato, indici negativi indicano invece che il valore target diminuisce con il crescere del valore del fattore correlato.

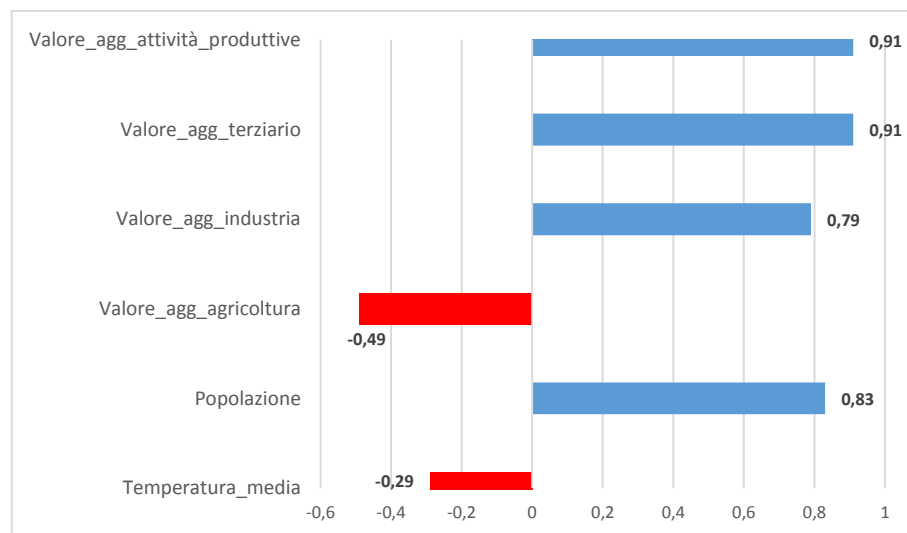


Figura 10.2 - Indici di correlazione per consumo in ktep nel residenziale.  
(■ valore negativo - ■ valori positivi)

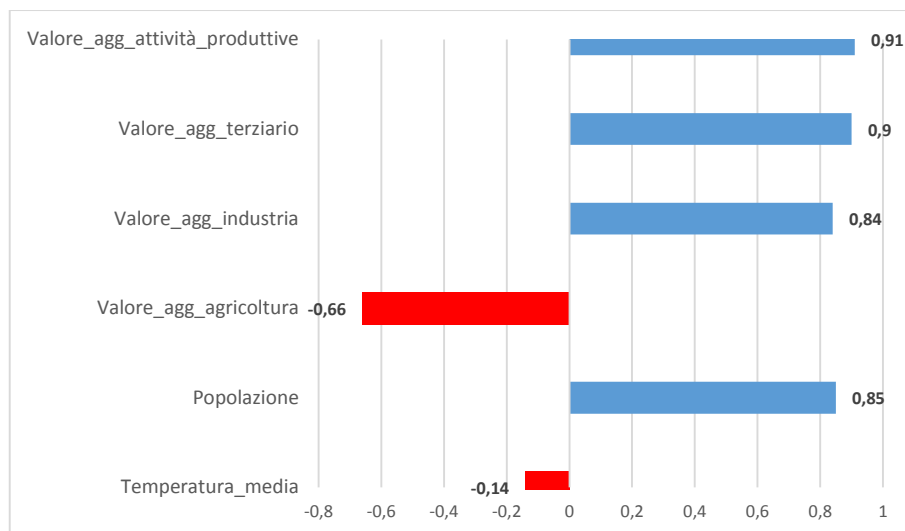


Figura 10.3 - Indici di correlazione per consumo in ktep nei servizi non vendibili.  
(■ valore negativo - ■ valori positivi)

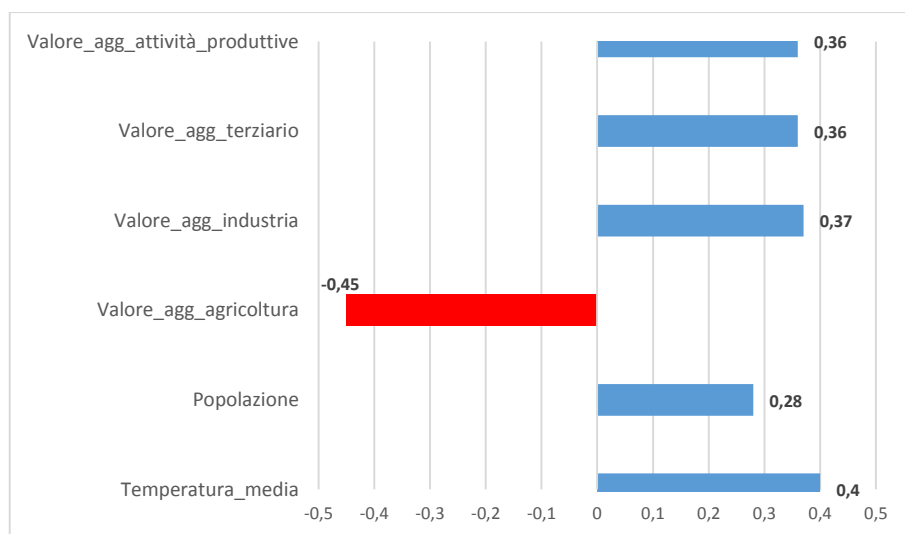


Figura 10.4 - Indici di correlazione per consumo in ktep in agricoltura.  
(■ valore negativo - ■ valori positivi)

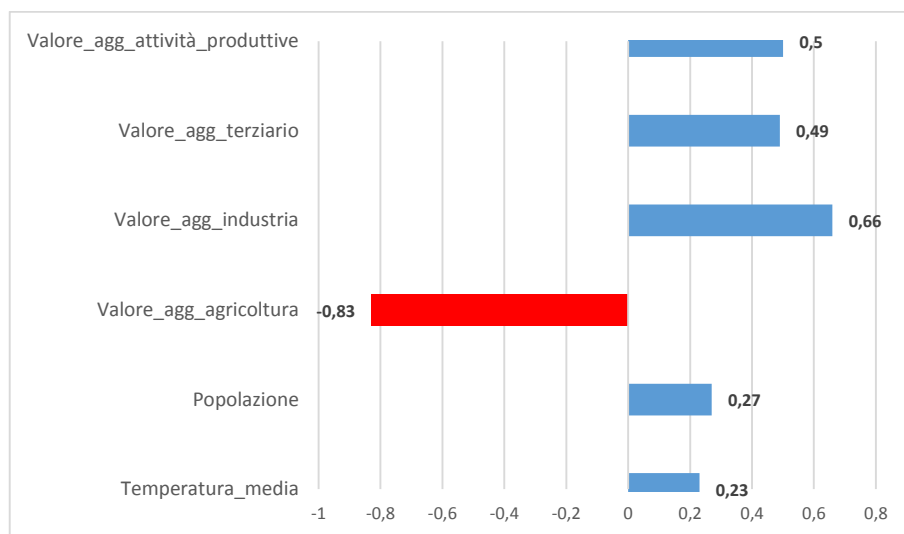


Figura 10.5 - Indici di correlazione per consumo in ktep nell'industria.

(■ valore negativo - ■ valori positivi)

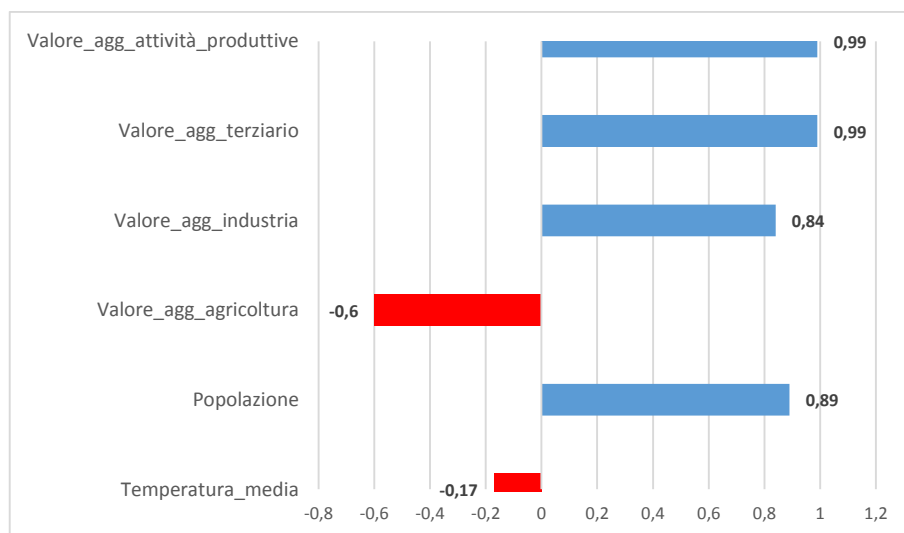


Figura 10.6 - Indici di correlazione per consumo in ktep nel terziario.

(■ valore negativo - ■ valori positivi)

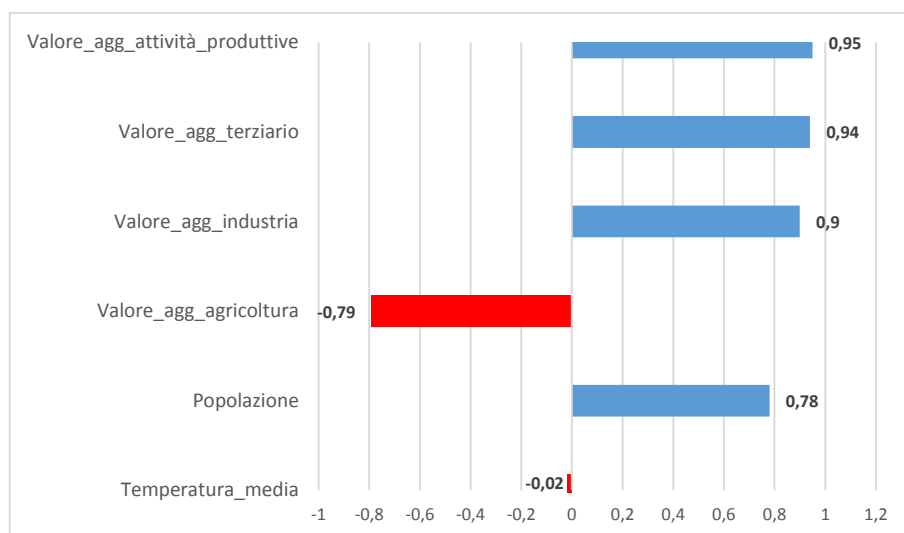


Figura 10.7 - Indici di correlazione per consumo in ktep nelle attività produttive.  
(■ valore negativo - ■ valori positivi)

### 10.1.3 Modello di previsione Neural Network “Consumi EE regionali”

Il modello sviluppato con il software GMDH Shell basato su GMDH Neural Network ha come scopo la definizione dei consumi di energia elettrica annuale per singolo settore ponendo come input della rete neurale i dati relativi alla popolazione, al valore aggiunto per singolo settore, ed alla temperatura media annuale. Nella Figura 10.8 si riporto lo schema della rete utilizzata.

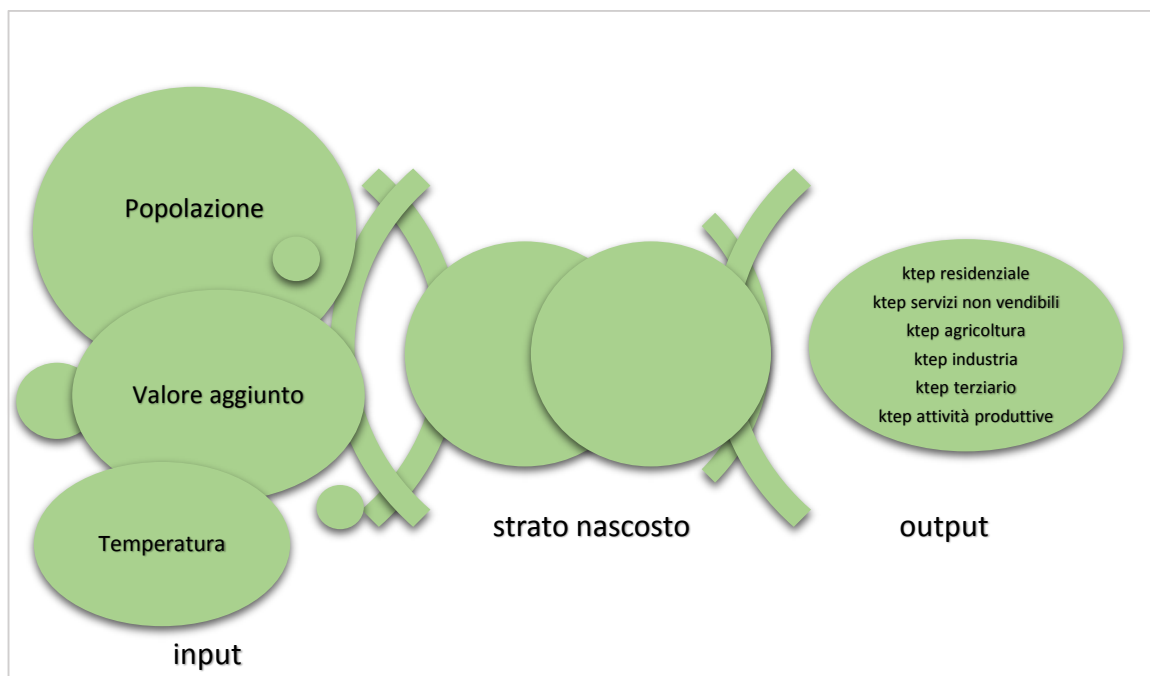


Figura 10.8 – Diagramma neurale - Previsione “Consumi EE regionali”.



Attraverso l'iterazione di più settaggi si è scelto quello che coniugasse basso errore e velocità computazionale. Sono state quindi definite le seguenti impostazioni nel pannello Solver (Tabella 10.9).

|                       |                       |
|-----------------------|-----------------------|
| Reorder observations  | Odd/even              |
| Validation strategy   | k-fold validation     |
| Number of folds       | 2                     |
| Validation criterion  | RMSE                  |
| Variables ranking     | No                    |
| Core algorithm        | GMDH neural network   |
| Neuron function       | $a+x_i+x_i*x_j+x_i^2$ |
| Max. number of layers | 5                     |
| Initial layer width   | 10                    |
| Time series mode      | -                     |

Tabella 10.9 – Settings solver – Modello di previsione “Consumi EE regionali”.

Si è quindi scelto come: criterio di validazione il RMSE, che rappresenta la deviazione standard del campione delle differenze tra i valori previsti ed i valori osservati; core algorithm (GMDH neural network) che rappresenta l'algoritmo di apprendimento statistico; la funzione neurale di attivazione del singolo neurone ( $a+x_i+x_i*x_j+x_i^2$ ) con la quale si ottiene il plot del model fit e della predictions; massimo numero di strati (5) come limite superiore per il numero di strati di rete creati dall'algoritmo; e come larghezza dello strato iniziale (10) che definisce quanti neuroni vengono aggiunti alla serie di ingresso ad ogni nuovo livello.

Avviando il processo di training e di validation si sono quindi ottenuti i seguenti risultati per i singoli target (Tabelle 10.10 – 10.15).

| Postprocessed results                     | Model fit              | Predictions |
|---|------------------------|-------------|
| Number of observations                    | 18                     | 5           |
| Max. negative error                       | -11,97                 | -40,97      |
| Max. positive error                       | 8,59                   | 5,46        |
| Normalized Mean absolute error (nMAE)     | 4,34 %                 | 17,3 %      |
| Normalized Root mean square error (nRMSE) | 5,69 %                 | 22,2 %      |
| Residual sum                              | $-4,55 \cdot 10^{-11}$ | -68,41      |
| Standard deviation of residuals           | 5,54                   | 16,81       |
| Coefficient of determination ( $R^2$ )    | 0,95                   | -0,10       |
| Correlation                               | 0,97                   | 0,90        |

Tabella 10.10 – Accuracy ktep residenziale  
Modello di previsione “Consumi EE regionali” con Neural Network.

| Postprocessed results                          | Model fit | Predictions |
|--|-----------|-------------|
| Number of observations                         | 18        | 5           |
| Max. negative error                            | -10,77    | -11,05      |
| Max. positive error                            | 7,30      | 5,46        |
| Normalized Mean absolute error (nMAE)          | 4,86 %    | 9,05 %      |
| Normalized Root mean square error (nRMSE)      | 6,50 %    | 9,36 %      |
| Residual sum                                   | 2,31 e-11 | -1,95       |
| Standard deviation of residuals                | 4,10      | 5,89        |
| Coefficient of determination (R <sup>2</sup> ) | 0,96      | 0,83        |
| Correlation                                    | 0,98      | 0,97        |

Tabella 10.11 – Accuracy ktep servizi non vendibili  
Modello di previsione “Consumi EE regionali” con Neural Network.

| Postprocessed results                          | Model fit             | Predictions |
|--|-----------------------|-------------|
| Number of observations                         | 18                    | 5           |
| Max. negative error                            | -5,41                 | -11,72      |
| Max. positive error                            | 2,48                  | 2,64        |
| Normalized Mean absolute error (nMAE)          | 11,83 %               | 38,30 %     |
| Normalized Root mean square error (nRMSE)      | 16,41%                | 53,16%      |
| Residual sum                                   | -6,18e <sup>-11</sup> | -15,52      |
| Standard deviation of residuals                | 1,78                  | 4,87        |
| Coefficient of determination (R <sup>2</sup> ) | 0,66                  | -20,61      |
| Correlation                                    | 0,81                  | 0,13        |

Tabella 10.12 – Accuracy ktep agricoltura  
Modello di previsione “Consumi EE regionali” con Neural Network.

| Postprocessed results                          | Model fit | Predictions |
|--|-----------|-------------|
| Number of observations                         | 18        | 5           |
| Max. negative error                            | -19,05    | -12,16      |
| Max. positive error                            | 18,67     | 10,23       |
| Normalized Mean absolute error (nMAE)          | 6,02 %    | 4,21 %      |
| Normalized Root mean square error (nRMSE)      | 7,14%     | 5,43%       |
| Residual sum                                   | 0         | 0,11        |
| Standard deviation of residuals                | 9,68      | 7,37        |
| Coefficient of determination (R <sup>2</sup> ) | 0,94      | 0,96        |
| Correlation                                    | 0,97      | 0,98        |

Tabella 10.13 – Accuracy ktep industria  
Modello di previsione “Consumi EE regionali” con Neural Network.

| Postprocessed results                          | Model fit            | Predictions |
|--|----------------------|-------------|
| Number of observations                         | 18                   | 5           |
| Max. negative error                            | -9,11                | -67,67      |
| Max. positive error                            | 6,75                 | 10,82       |
| Normalized Mean absolute error (nMAE)          | 1,08 %               | 7,28 %      |
| Normalized Root mean square error (nRMSE)      | 1,48%                | 13,3%       |
| Residual sum                                   | 2,11e <sup>-12</sup> | -53,25      |
| Standard deviation of residuals                | 3,43                 | 28,81       |
| Coefficient of determination (R <sup>2</sup> ) | 0,99                 | 0,85        |
| Correlation                                    | 0,99                 | 0,94        |

Tabella 10.14 – Accuracy ktep terziario  
Modello di previsione “Consumi EE regionali” con Neural Network.

| Postprocessed results                     | Model fit             | Predictions |
|---|-----------------------|-------------|
| Number of observations                    | 18                    | 5           |
| Max. negative error                       | -36,17                | -115,81     |
| Max. positive error                       | 27,67                 | 0           |
| Normalized Mean absolute error (nMAE)     | 4,20 %                | 23,47 %     |
| Normalized Root mean square error (nRMSE) | 5,42%                 | 28,9%       |
| Residual sum                              | -4,50e <sup>-11</sup> | -336,00     |
| Standard deviation of residuals           | 15,54                 | 48,44       |
| Coefficient of determination ( $R^2$ )    | 0,97                  | 0,23        |
| Correlation                               | 0,98                  | 0,95        |

Tabella 10.15 – Accuracy ktep attività produttive  
Modello di previsione “Consumi EE regionali” con Neural Network.

Il calcolo del MAE e del RMSE in forma normalizzata permette di superare l'errore che si genera con il MAPE ed il RMSPE per via della prossimità al valore 0 dei dati di target.

Le Figure 10.9-10.14 rappresentano il plot del model fit e della previsione. In grigio sono rappresentati i dati reali del DataSet, in blu il risultato del model fit, in rosso il risultato della previsione e la banda rosa rappresenta la banda di confidenza della previsione.

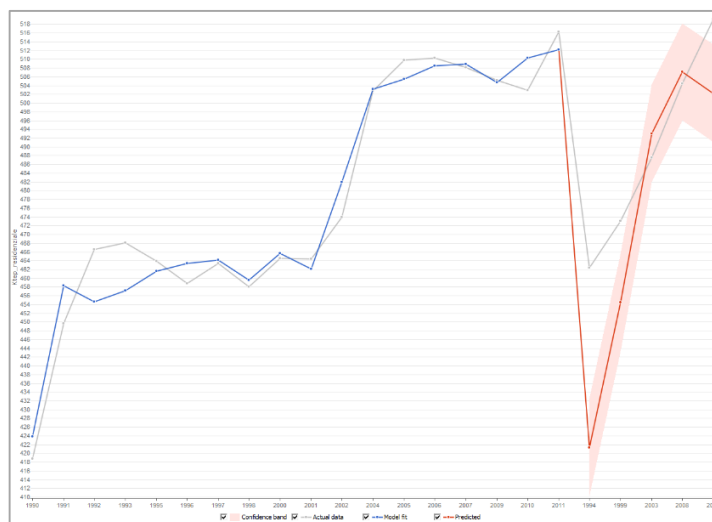


Figura 10.9 – Plot model fit e previsione ktep residenziale  
Modello di previsione “Consumi EE regionali” con Neural Network.

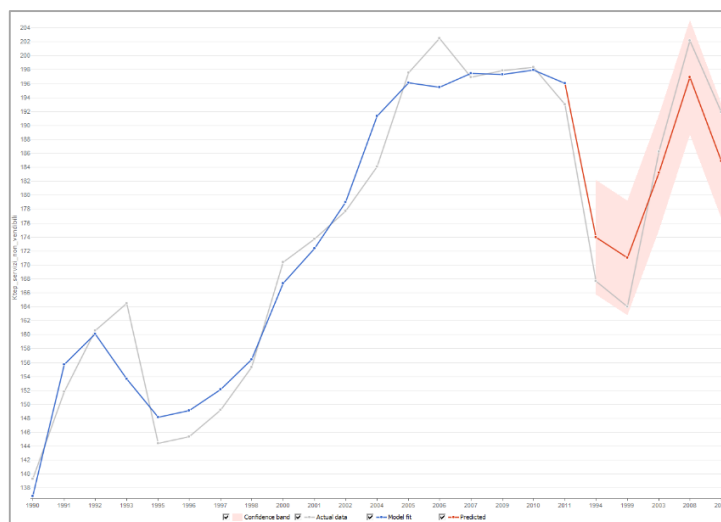


Figura 10.10 – Plot model fit e previsione ktep servizi non vendibili  
Modello di previsione “Consumi EE regionali” con Neural Network.

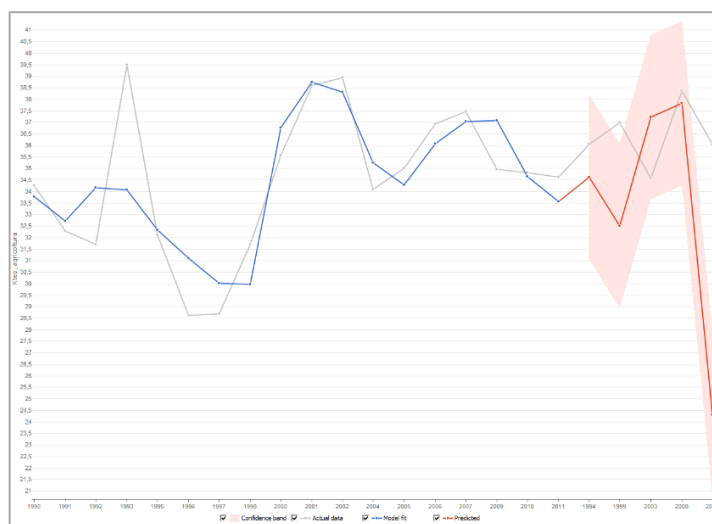


Figura 10.11 – Plot model fit e previsione ktep agricoltura  
Modello di previsione “Consumi EE regionali” con Neural Network.

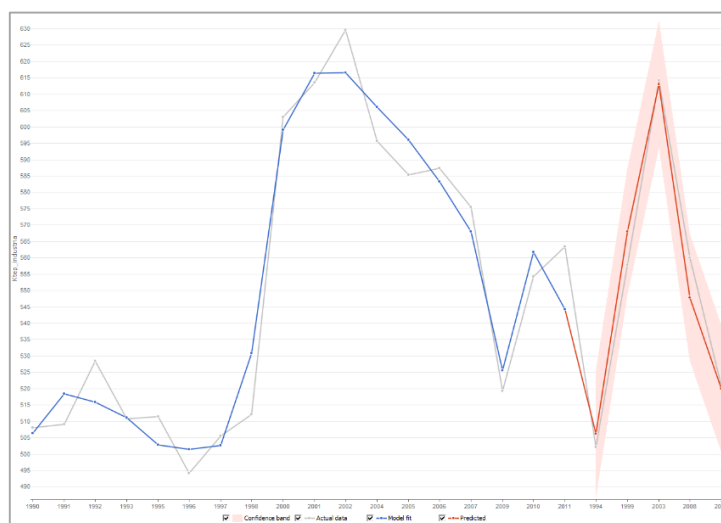


Figura 10.12 – Plot model fit e previsione ktep industria  
Modello di previsione “Consumi EE regionali” con Neural Network.

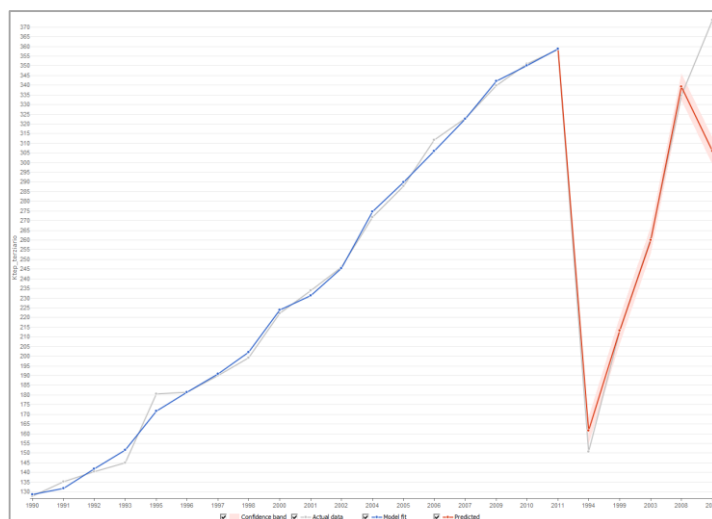


Figura 10.13 – Plot model fit e previsione ktep terziario  
Modello di previsione “Consumi EE regionali” con Neural Network.

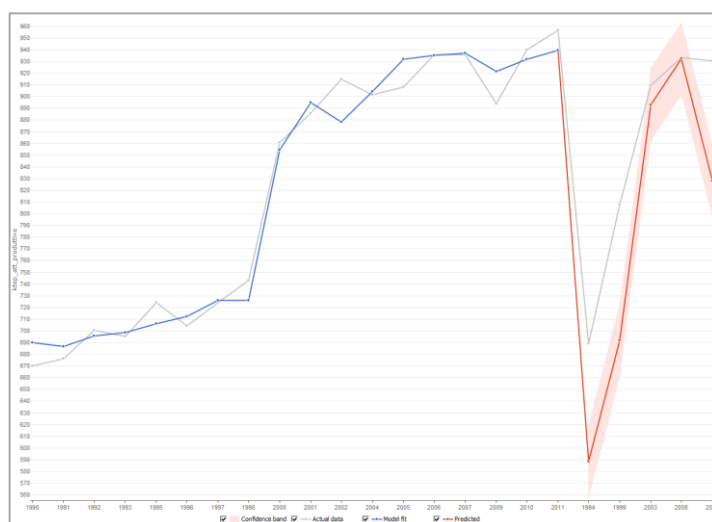


Figura 10.14 – Plot model fit e previsione ktep attività produttive  
Modello di previsione “Consumi EE regionali” con Neural Network.

Di seguito si riportano le funzioni generate per la predizione.

### Consumi residenziale

$$Y1 = 243.289 - N16 * N2 * 0.000735403 + N2^2 * 0.00175956$$

$$N2 = -32213.2 + \text{Popolazione} * 0.00648096 - \text{Popolazione} * N6 * 1.26882e-05 + N6 * 64.0644$$

$$N6 = 255.051 + N14 * N9 * 0.331395 - N14^2 * 0.165434 - N9^2 * 0.164976$$

$$N9 = 233.859 + N13 * N14 * 0.00106573$$

$$N13 = 566.475 + \text{Temperatura\_media} * \text{Valore\_agg\_terziario} * 1.24395e-10 - \text{Temperatura\_media}^2 * 0.640761$$

$$N14 = 3356.21 - \text{Valore\_agg\_agricoltura} * 2.14925e-06 + \text{Valore\_agg\_agricoltura} * \text{Valore\_agg\_attività\_produttive} * 9.8467e-19 + \text{Valore\_agg\_agricoltura}^2 * 3.78368e-16$$

$$N16 = 521.588 + \text{Valore\_agg\_agricoltura} * 1.2475e-07 - \text{Valore\_agg\_industria} * 1.25501e-07 + \text{Valore\_agg\_industria}^2 * 8.38388e-18$$

### Consumi attività non vendibili

$$Y1 = -140.247 - N11 * N2 * 0.00485841 + N2^2 * 2.66307$$

$$N2 = -167.767 + \text{Popolazione} * N5 * 6.21994e-07 - N5^2 * 0.00648946$$

$$N5 = 85.7366 - N16 * N8 * 0.128041 + N16^2 * 0.0625021 + N8^2 * 0.0683544$$

$$N8 = -481.3 - \text{Temperatura\_media} * N13 * 0.337068 + \text{Temperatura\_media}^2 * 1.67954 + N13 * 6.70221$$

$$N13 = 2563.47 - \text{Valore\_agg\_agricoltura} * 1.71778e-06 + \text{Valore\_agg\_agricoltura}^2 * 3.00882e-16 + \text{Valore\_agg\_attività\_produttive}^2 * 1.39669e-20$$

$$N16 = 3154.15 - \text{Temperatura\_media} * 349.321 + \text{Temperatura\_media} * \text{Valore\_agg\_terziario} * 1.01743e-10 + \text{Temperatura\_media}^2 * 9.94978$$

$$N11 = 2613.37 - \text{Valore\_agg\_agricoltura} * 1.74335e-06 + \text{Valore\_agg\_agricoltura}^2 * 3.0477e-16 + \text{Valore\_agg\_terziario}^2 * 1.87408e-20$$

### **Consumi agricoltura**

$$Y1 = 2.96457 + \text{Temperatura\_media} * N5 * 0.053813$$

$$\begin{aligned} N5 &= 21.5108 + N13 * N7 * 0.636573 - N13^2 * 0.323258 - N7^2 * 0.301751 \\ N7 &= 60676.2 - \text{Popolazione} * 0.0242462 + \text{Popolazione}^2 * 2.42153e-09 + N11 * 1.41467 \\ N11 &= 466.731 - \text{Valore\_agg\_agricoltura} * 2.96271e-07 + \text{Valore\_agg\_agricoltura}^2 * 5.0484e-17 \\ N13 &= 20.6701 + \text{Temperatura\_media} * \text{Valore\_agg\_industria} * 7.86941e-11 \end{aligned}$$

### **Consumi industria**

$$Y1 = 5.85892 + \text{Popolazione} * N3 * 1.981e-07$$

$$\begin{aligned} N3 &= 1522.62 + N17 * N4 * 0.00983638 - N17^2 * 0.00483512 - N4^2 * 4.52686 \\ N4 &= 284.86 + N17 * N7 * 0.0235177 - N17^2 * 0.0119213 - N7^2 * 0.0106978 \\ N7 &= -7118.59 + \text{Valore\_agg\_industria} * 6.45795e-07 - \text{Valore\_agg\_industria} * N13 * 1.24968e-09 + N13 * 14.7887 \\ N13 &= 5465.83 - \text{Valore\_agg\_agricoltura} * 3.28264e-06 + \text{Valore\_agg\_agricoltura}^2 * 5.43552e-16 \\ N17 &= 26.8225 + \text{Valore\_agg\_industria} * 6.33036e-08 - \text{Valore\_agg\_industria} * \text{Valore\_agg\_attività\_produttive} * 2.08543e-19 \end{aligned}$$

### **Consumi terziario**

$$Y1 = 0.831193 + N9 * N3 * 0.000679206 + N3 * 0.992206 - N3^2 * 0.000662787$$

$$\begin{aligned} N3 &= -0.683676 + N7 * 1.00638 + N7 * N4 * 0.00471632 - N7^2 * 0.00472967 \\ N4 &= 5.95922 - N15 * 1.20732 + N11 * 2.15337 + N11^2 * 0.000110084 \\ N11 &= 37.2848 - \text{Valore\_agg\_agricoltura}^2 * 2.47614e-18 + N14 * 0.868318 + N14^2 * 0.000219144 \\ N14 &= -2108.98 + \text{Popolazione} * 0.000420641 + \text{Valore\_agg\_terziario} * 4.81224e-09 + \text{Valore\_agg\_terziario}^2 * 1.10126e-20 \\ N15 &= -48547.1 + \text{Popolazione} * 0.0192241 - \text{Popolazione}^2 * 1.90066e-09 + \text{Valore\_agg\_attività\_produttive}^2 * 5.05556e-20 \\ N7 &= -798.58 + \text{Popolazione} * N13 * 2.56101e-07 + \text{Popolazione}^2 * 3.14982e-11 - N13^2 * 0.000866926 \\ N13 &= 172.004 - \text{Valore\_agg\_industria} * 2.82013e-08 + \text{Valore\_agg\_industria} * \text{Valore\_agg\_terziario} * 4.40681e-19 + \text{Valore\_agg\_terziario}^3 * 1.1486e-09 \\ N9 &= -1.74264 - N12 * N16 * 0.00698105 + N12^2 * 0.0069354 + N16^2 * 1.01914 \\ N16 &= 49.2541 - \text{Temperatura\_media} * \text{Valore\_agg\_attività\_produttive} * 2.0811e-11 + \text{Valore\_agg\_attività\_produttive}^2 * 5.75871e-20 \\ N12 &= 125.96 + \text{Valore\_agg\_industria} * \text{Valore\_agg\_attività\_produttive} * 2.48396e-18 - \text{Valore\_agg\_industria}^2 * 8.57382e-18 - \text{Valore\_agg\_attività\_produttive}^2 * 1.4015e-1 \end{aligned}$$

### **Consumi attività produttive**

$$Y1 = 406.905 + N12 * N2 * 0.0409202 - N12^2 * 0.0206026 - N2^2 * 0.0197013$$

$$\begin{aligned} N2 &= 11380.5 - N7 * 27.765 - N7 * N6 * 0.0349894 + N7^2 * 0.052852 \\ N6 &= 396.59 + N7^2 * 0.000619594 \\ N7 &= -65805 + \text{Popolazione} * 0.0132407 - \text{Popolazione} * N14 * 1.40695e-05 + N14 * 70.9191 \\ N14 &= 20475.7 - \text{Temperatura\_media} * 2368.48 + \text{Temperatura\_media}^2 * 70.5094 + \text{Valore\_agg\_terziario}^2 * 9.84827e-20 \\ N12 &= 1413.41 - \text{Valore\_agg\_agricoltura} * 2.59652e-07 + \text{Valore\_agg\_terziario}^2 * 6.24442e-2 \end{aligned}$$

#### **10.1.4 Modello di previsione serie storiche “Consumi EE regionali”**

Il modello sviluppato con il software GMDH Shell basato su *Stepwise mixed selection* di GMDH ha come scopo la definizione dei consumi di energia elettrica annuale per singolo settore ponendo come input i dati storici annuali della variabile di output dall'anno 1990 all'anno 2012.

Dopo diverse iterazioni con più settaggi si è scelto quello che coniugasse basso errore e velocità computazionale. Sono state quindi definite le seguenti impostazioni nel pannello Solver (Tabella 10.16).

|                           |                          |
|---------------------------|--------------------------|
| Reorder observations      | Pseud-random             |
| Validation strategy       | k-fold validation        |
| Number of folds           | 2                        |
| Validation criterion      | RMSE                     |
| Variables ranking         | No                       |
| Drop variables after rank | -                        |
| Core algorithm            | Stepwise mixed selection |
| Neuron function           | $x_i * x_j, x_j^2$       |
| Limit model complexitt to | 100                      |
| Return set of best models | 100                      |
| Time series mode          | -                        |

Tabella 10.16 – Settings solver – Modello di previsione “Consumi EE regionali”.

Avviando il processo di traning con la previsione dei successivi 5 anni, si sono ottenuti i seguenti risultati (Tabelle 10.17 – 10.22).

| Postprocessed results                     | Model fit      |
|---|----------------|
| Number of observations                    | 18             |
| Max. negative error                       | -2,19          |
| Max. positive error                       | 2,74           |
| Normalized Mean absolute error (nMAE)     | 1,90 %         |
| Normalized Root mean square error (nRMSE) | 2,08 %         |
| Residual sum                              | $-4,02e^{-11}$ |
| Standard deviation of residuals           | 1,90           |
| Coefficient of determination ( $R^2$ )    | 0,98           |
| Correlation                               | 0,99           |

Tabella 10.17 – Accuracy ktep residenziale– Modello di previsione “Consumi EE regionali”.

| Postprocessed results                     | Model fit      |
|---|----------------|
| Number of observations                    | 18             |
| Max. negative error                       | -3,27          |
| Max. positive error                       | 5,24           |
| Normalized Mean absolute error (nMAE)     | 3,43 %         |
| Normalized Root mean square error (nRMSE) | 4,17 %         |
| Residual sum                              | $-5,40e^{-12}$ |
| Standard deviation of residuals           | 2,63           |
| Coefficient of determination ( $R^2$ )    | 0,94           |
| Correlation                               | 0,97           |

Tabella 10.18 – Accuracy ktep servizi non vendibili– Modello di previsione “Consumi EE regionali”.

| Postprocessed results                          | Model fit            |
|--|----------------------|
| Number of observations                         | 18                   |
| Max. negative error                            | -1,31                |
| Max. positive error                            | 1,23                 |
| Normalized Mean absolute error (nMAE)          | 6,61 %               |
| Normalized Root mean square error (nRMSE)      | 7,30 %               |
| Residual sum                                   | 1,45e <sup>-12</sup> |
| Standard deviation of residuals                | 0,79                 |
| Coefficient of determination (R <sup>2</sup> ) | 0,79                 |
| Correlation                                    | 0,93                 |

Tabella 10.19 – Accuracy ktep agricoltura– Modello di previsione “Consumi EE regionali”.

| Postprocessed results                          | Model fit             |
|--|-----------------------|
| Number of observations                         | 18                    |
| Max. negative error                            | -10,58                |
| Max. positive error                            | 7,18                  |
| Normalized Mean absolute error (nMAE)          | 3,30 %                |
| Normalized Root mean square error (nRMSE)      | 4,03 %                |
| Residual sum                                   | -9,03e <sup>-10</sup> |
| Standard deviation of residuals                | 5,47                  |
| Coefficient of determination (R <sup>2</sup> ) | 0,89                  |
| Correlation                                    | 0,94                  |

Tabella 10.20 – Accuracy ktep industria– Modello di previsione “Consumi EE regionali”.

| Postprocessed results                          | Model fit            |
|--|----------------------|
| Number of observations                         | 18                   |
| Max. negative error                            | -3,31                |
| Max. positive error                            | 2,85                 |
| Normalized Mean absolute error (nMAE)          | 0,77%                |
| Normalized Root mean square error (nRMSE)      | 0,95 %               |
| Residual sum                                   | 3,43e <sup>-12</sup> |
| Standard deviation of residuals                | 1,85                 |
| Coefficient of determination (R <sup>2</sup> ) | 0,99                 |
| Correlation                                    | 0,99                 |

Tabella 10.21 – Accuracy ktep terziario– Modello di previsione “Consumi EE regionali”.

| Postprocessed results                          | Model fit             |
|--|-----------------------|
| Number of observations                         | 18                    |
| Max. negative error                            | -14,59                |
| Max. positive error                            | 13,65                 |
| Normalized Mean absolute error (nMAE)          | 2,65%                 |
| Normalized Root mean square error (nRMSE)      | 3,23%                 |
| Residual sum                                   | -3,42e <sup>-11</sup> |
| Standard deviation of residuals                | 8,59                  |
| Coefficient of determination (R <sup>2</sup> ) | 0,86                  |
| Correlation                                    | 0,93                  |

Tabella 10.22 – Accuracy ktep attività produttive– Modello di previsione “Consumi EE regionali”.



Le figure 10.15 – 10.20 rappresentano il plot del model fit e della predizione dei singoli target. In grigio sono rappresentati i dati reali del DataSet, in blu il risultato del model fit, in rosso il risultato della predizione e la banda rosa rappresenta la banda di confidenza della predizione.

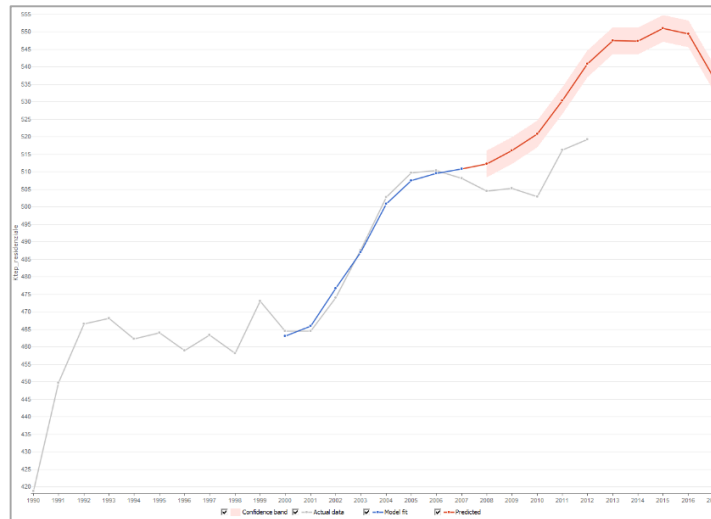


Figura 10.15 – Plot model fit e previsione ktep residenziale–  
Modello di previsione “Consumi EE regionali”.

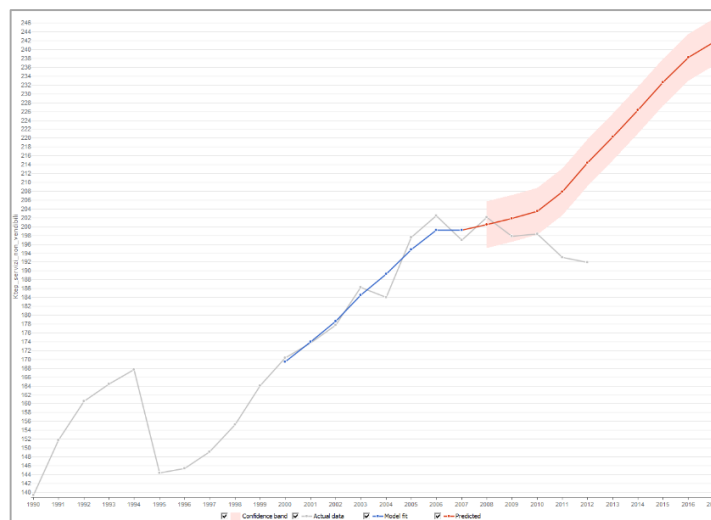


Figura 10.16 – Plot model fit e previsione ktep servizi non vendibili–  
Modello di previsione “Consumi EE regionali”.

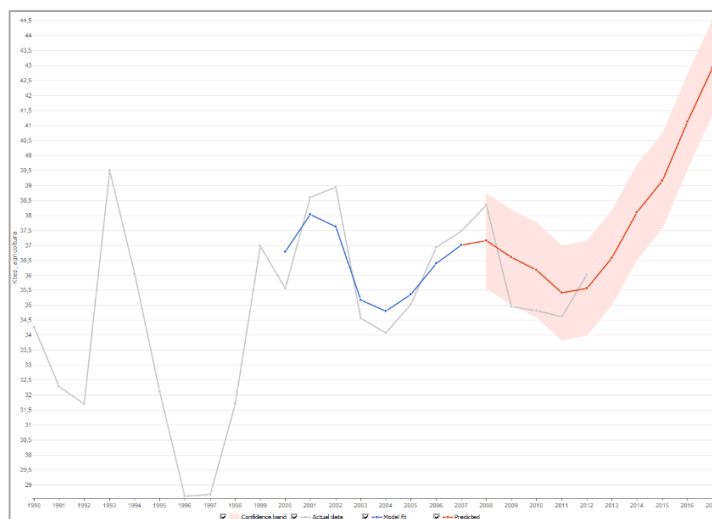


Figura 10.17 – Plot model fit e previsione ktep agricoltura–  
Modello di previsione “Consumi EE regionali”.

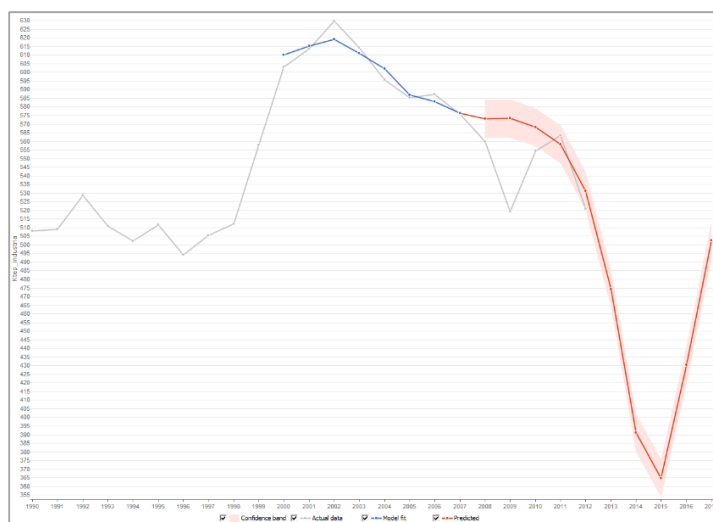


Figura 10.18 – Plot model fit e previsione ktep industria–  
Modello di previsione “Consumi EE regionali”.

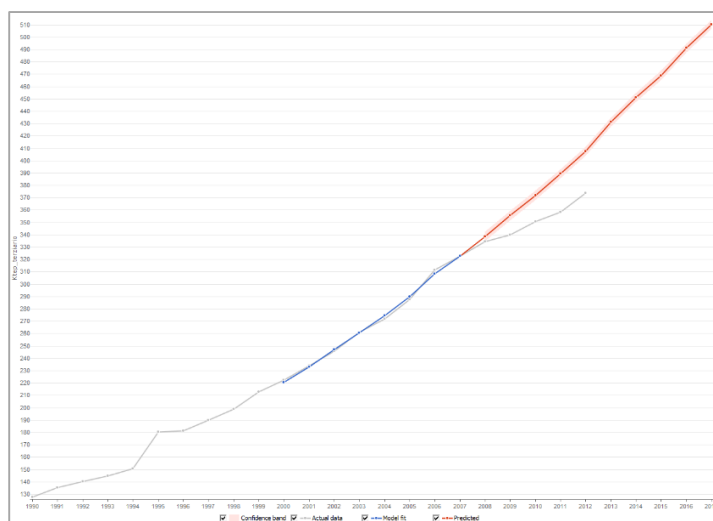


Figura 10.19 – Plot model fit e previsione ktep terziario–  
Modello di previsione “Consumi EE regionali”.

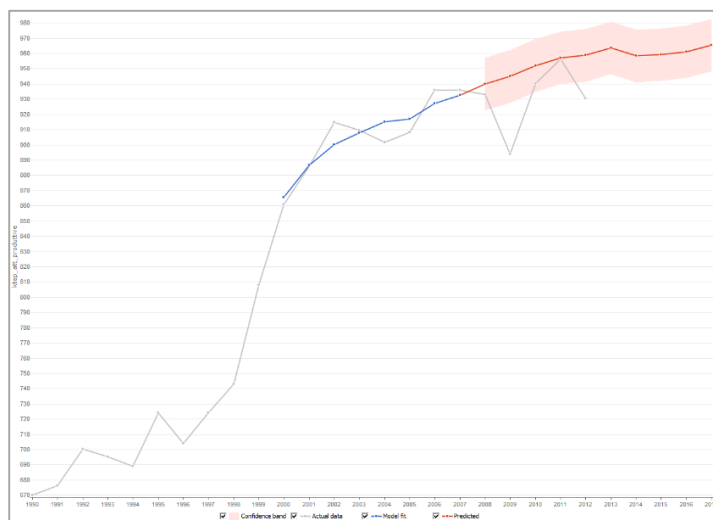


Figura 10.20 – Plot model fit e previsione ktep attività produttive  
Modello di previsione “Consumi EE regionali”.

## 10.2 Scala provinciale

I dati forniti su scala regionale sono anche forniti a scala provinciale per gli ultimi sei anni. Ciò comporta il poter disporre di un numero maggiore di valori nel DataSET che passa da 23 a 54 righe, ma perdendo la caratteristica di serie storica. Pertanto sarà di seguito sviluppato solo il modello di previsione regressivo NN.

Il data base relativo alle nove province riporta sempre gli indicatori energetici, ambientali, demografici ed economici. Per le stesse considerazioni fatte per il caso studio su scala regionale, anche per questo caso studio, si è svolta un’analisi predittiva dei consumi di energia elettrica per singolo settore (residenziale, servizi non vendibili, agricoltura industria, terziario ed attività produttive). Come variabili di input sono stati scelti i valori aggiunti dei singoli settori, il numero della popolazione e la temperatura media annuale.

### 10.2.1 *Il DataSet*

Tramite sempre i dati provenienti dal Ministero dello Sviluppo Economico[36-37], TERNA Spa [38], GSE Spa [39-42], ISPRA [43], ISTAT [44] ed Istituto Guglielmo Tagliacarne [45] è stato possibile la costruzione del Database provinciale che si riporta nelle Tabelle 10.23-10.26

| T media [°C] | Popolazione | Valore_agg_ Agricoltura [€] | Valore_agg_ Industria [€] | Valore_agg_ Terziario [€] | Co <sub>2</sub> _fatt [tCO <sub>2</sub> /tep] |
|--------------|-------------|-----------------------------|---------------------------|---------------------------|---|
| 17,59        | 452294      | 301121865                   | 730484339                 | 4522532010                | 5,51  |
| 17,71        | 272870      | 148912752                   | 828324278                 | 2663052976                | 5,51  |
| 16,65        | 1074498     | 392542641                   | 2566192886                | 12150640685               | 5,51  |
| 16,33        | 172048      | 140124979                   | 325432177                 | 1695421659                | 5,51  |
| 15,43        | 668365      | 224810287                   | 15180063                  | 7668641491                | 5,51  |
| 16,72        | 1248262     | 326580367                   | 2418256834                | 14864922900               | 5,51  |

|       |         |            |             |              |      |
|-------|---------|------------|-------------|--------------|------|
| 17,33 | 299596  | 418050436  | 785811590   | 3147976162   | 5,51 |
| 18,13 | 403028  | 346167466  | 1262932553  | 4240498829   | 5,51 |
| 17,65 | 431982  | 332903354  | 853845595   | 4571392070   | 5,51 |
| 17,58 | 453142  | 309373134  | 743843695   | 4738915248   | 5,34 |
| 17,83 | 272637  | 149750460  | 836946361   | 2744438323   | 5,34 |
| 17,02 | 1078442 | 405852766  | 2589588838  | 12432938219  | 5,34 |
| 16,30 | 171017  | 139901201  | 339372172   | 1759007842   | 5,34 |
| 15,46 | 669548  | 222517036  | 1545224039  | 7865591928   | 5,34 |
| 16,81 | 1250730 | 319563337  | 2436689170  | 15370437072  | 5,34 |
| 17,55 | 300462  | 409199442  | 813564877   | 3253064551   | 5,34 |
| 18,37 | 404400  | 363041151  | 1228689468  | 4405799260   | 5,34 |
| 17,96 | 433354  | 315260961  | 853953005   | 4724101761   | 5,34 |
| 17,83 | 453990  | 317624403  | 757203051   | 4955298486   | 5,21 |
| 17,93 | 272404  | 150588168  | 845568444   | 2825823669   | 5,21 |
| 16,86 | 1082386 | 419162891  | 2612984789  | 12715235754  | 5,21 |
| 16,02 | 169987  | 139677422  | 353312168   | 1822594025   | 5,21 |
| 15,20 | 670731  | 220223786  | 1572441689  | 8062542366   | 5,21 |
| 16,93 | 1253198 | 312546306  | 2455121507  | 15875951244  | 5,21 |
| 18,98 | 301329  | 400348449  | 841318164   | 3358152939   | 5,21 |
| 18,21 | 405772  | 379914835  | 1194446382  | 4571099690   | 5,21 |
| 17,78 | 434726  | 305831199  | 854060414   | 4876811453   | 5,21 |
| 17,81 | 454838  | 325875672  | 770562408   | 5171681724   | 4,80 |
| 17,52 | 272171  | 151425876  | 854190528   | 2907209016   | 4,80 |
| 16,57 | 1086330 | 432473016  | 2636380740  | 12997533288  | 4,80 |
| 15,63 | 173262  | 139453644  | 367252164   | 1886180208   | 4,80 |
| 15,66 | 654206  | 217930536  | 1599659340  | 8259492804   | 4,80 |
| 16,99 | 1245387 | 305529276  | 2473553844  | 16381465416  | 4,80 |
| 17,90 | 315007  | 391497456  | 869071452   | 3463241328   | 4,80 |
| 18,09 | 403098  | 396788520  | 1160203296  | 4736400120   | 4,80 |
| 18,00 | 436098  | 296401438  | 854167824   | 5029521144   | 4,80 |
| 17,27 | 454593  | 346937836  | 746281204   | 4992840862   | 4,60 |
| 17,66 | 272052  | 179712938  | 746095264   | 2866604508   | 4,60 |
| 16,70 | 1087682 | 424736508  | 2552190370  | 13205766644  | 4,60 |
| 15,75 | 173009  | 140226822  | 341626082   | 1869590104   | 4,60 |
| 14,87 | 653810  | 278465268  | 1551829670  | 8569746402   | 4,60 |
| 16,28 | 1246094 | 340764638  | 2448776922  | 16780232708  | 4,60 |
| 17,48 | 316113  | 396748728  | 814035726   | 3610120664   | 4,60 |
| 17,93 | 403356  | 338394260  | 1149101648  | 4714700060   | 4,60 |
| 17,57 | 436283  | 273911434  | 882083912   | 5012760572   | 4,60 |
| 17,44 | 454002  | 368000000* | 722000000*  | 4814000000*  | 4,57 |
| 17,79 | 271279  | 208000000* | 638000000*  | 2826000000*  | 4,57 |
| 16,43 | 1090101 | 417000000* | 2468000000* | 13414000000* | 4,57 |
| 15,73 | 172485  | 141000000* | 316000000*  | 1853000000*  | 4,57 |
| 14,78 | 653737  | 339000000* | 1504000000* | 8880000000*  | 4,57 |
| 16,39 | 1249577 | 376000000* | 2424000000* | 17179000000* | 4,57 |
| 17,24 | 318549  | 402000000* | 759000000*  | 3757000000*  | 4,57 |
| 17,50 | 404271  | 280000000* | 1138000000* | 4693000000*  | 4,57 |
| 17,73 | 436624  | 308000000* | 910000000*  | 4996000000*  | 4,57 |

Tabella 10.23 – Valori climatici, economici ed ambientali DataSet Provincie Sicilia.

(\* dato disaggregato a partire dal dato totale Sicilia)

| Residenziale<br>[ktep] | Agricoltura<br>[ktep] | Industria<br>[ktep] | Terziario<br>[ktep] |
|------------------------|-----------------------|---------------------|---------------------|
| 44,50                  | 2,2446                | 20,8722             | 31,93               |
| 25,54                  | 1,3244                | 89,7066             | 19,78               |
| 100,77                 | 8,4624                | 93,7056             | 99,62               |
| 14,69                  | 1,1524                | 5,8738              | 11,83               |
| 70,20                  | 1,9866                | 88,5628             | 65,79               |
| 132,06                 | 2,5628                | 45,2532             | 108,60              |
| 32,22                  | 9,0214                | 43,688              | 26,72               |
| 42,97                  | 7,783                 | 238,5554            | 39,16               |
| 47,39                  | 2,3908                | 23,9854             | 35,89               |
| 43,65                  | 2,3392                | 22,36               | 32,95               |
| 25,46                  | 1,4964                | 92,1232             | 18,93               |
| 105,57                 | 9,116                 | 94,5398             | 105,96              |
| 14,88                  | 1,075                 | 5,031               | 12,23               |
| 68,22                  | 1,9522                | 83,9188             | 64,65               |
| 128,24                 | 2,5972                | 46,3368             | 107,54              |
| 32,24                  | 9,073                 | 45,2102             | 28,32               |
| 42,17                  | 7,4734                | 227,3496            | 37,07               |
| 47,70                  | 2,3478                | 22,9878             | 35,97               |
| 42,81                  | 2,365                 | 22,8072             | 32,34               |
| 25,26                  | 1,505                 | 81,9752             | 20,71               |
| 104,38                 | 8,9784                | 95,7524             | 109,21              |
| 14,71                  | 1,075                 | 5,934               | 12,23               |
| 68,28                  | 2,1586                | 81,1238             | 68,27               |
| 126,92                 | 2,6144                | 44,548              | 112,00              |
| 32,40                  | 9,6922                | 44,1352             | 30,30               |
| 42,70                  | 7,4562                | 224,4342            | 37,27               |
| 47,02                  | 2,5026                | 23,0738             | 38,17               |
| 43,28                  | 2,0984                | 19,9004             | 33,08               |
| 25,43                  | 1,3072                | 76,3336             | 23,31               |
| 104,07                 | 7,9894                | 85,441              | 110,22              |
| 14,90                  | 0,9632                | 5,8996              | 12,19               |
| 68,39                  | 1,8748                | 82,0096             | 68,87               |
| 127,60                 | 2,5886                | 40,6006             | 112,73              |
| 32,67                  | 8,9096                | 41,9164             | 29,93               |
| 42,71                  | 6,9316                | 204,0264            | 38,34               |
| 46,19                  | 2,2962                | 22,2224             | 37,12               |
| 43,76                  | 2,1242                | 19,1608             | 34,89               |
| 25,38                  | 1,3416                | 79,464              | 21,90               |
| 105,11                 | 8,1614                | 87,9178             | 114,13              |
| 14,88                  | 1,0148                | 5,9856              | 12,40               |
| 66,96                  | 1,8748                | 83,291              | 68,71               |
| 126,03                 | 2,5284                | 43,1806             | 116,70              |
| 32,46                  | 8,815                 | 45,0726             | 30,38               |
| 41,81                  | 6,7166                | 228,158             | 38,49               |
| 46,57                  | 2,2446                | 23,3146             | 38,75               |
| 44,58                  | 2,1758                | 17,3806             | 35,08               |
| 25,94                  | 1,376                 | 75,7832             | 22,45               |
| 108,98                 | 7,9464                | 92,4758             | 115,87              |
| 15,05                  | 1,0148                | 4,7902              | 14,14               |
| 67,34                  | 1,892                 | 86,9804             | 67,86               |
| 129,30                 | 2,6316                | 41,7272             | 117,36              |
| 33,47                  | 8,8408                | 41,5294             | 31,16               |
| 43,96                  | 6,4156                | 237,7728            | 39,35               |
| 47,61                  | 2,322                 | 21,543              | 39,35               |

Tabella 10.24 – Consumi energetici DataSet Provincie Sicilia.

| Residenziale<br>[ktep/p,ne] | Servizi_non_vendibili<br>[ktep/Mln,€] | Agricoltura<br>[ktep/Mln,€] | Industria<br>[ktep/Mln,€] | Terziario<br>[ktep/Mln,€] |
|-----------------------------|---------------------------------------|-----------------------------|---------------------------|---------------------------|
| 0,10                        | 7,45                                  | 4,62                        | 7,06                      | 19,13                     |
| 0,09                        | 8,89                                  | 33,69                       | 7,43                      | 50,01                     |
| 0,09                        | 21,56                                 | 7,71                        | 8,20                      | 37,47                     |
| 0,09                        | 8,22                                  | 3,46                        | 6,98                      | 18,67                     |
| 0,11                        | 8,84                                  | 11,55                       | 8,58                      | 28,96                     |
| 0,11                        | 7,85                                  | 3,04                        | 7,31                      | 18,20                     |
| 0,11                        | 21,58                                 | 13,88                       | 8,49                      | 43,95                     |
| 0,11                        | 22,48                                 | 56,26                       | 9,24                      | 87,98                     |
| 0,11                        | 7,18                                  | 5,25                        | 7,85                      | 20,28                     |
| 0,10                        | 7,56                                  | 4,72                        | 6,95                      | 19,23                     |
| 0,09                        | 9,99                                  | 33,57                       | 6,90                      | 50,46                     |
| 0,10                        | 22,46                                 | 7,60                        | 8,52                      | 38,59                     |
| 0,09                        | 7,68                                  | 2,86                        | 6,95                      | 17,50                     |
| 0,10                        | 8,77                                  | 10,67                       | 8,22                      | 27,66                     |
| 0,10                        | 8,13                                  | 3,01                        | 7,00                      | 18,14                     |
| 0,11                        | 22,17                                 | 13,90                       | 8,71                      | 44,78                     |
| 0,10                        | 20,59                                 | 51,60                       | 8,41                      | 80,60                     |
| 0,11                        | 7,45                                  | 4,87                        | 7,61                      | 19,93                     |
| 0,09                        | 7,45                                  | 4,60                        | 6,53                      | 18,58                     |
| 0,09                        | 9,99                                  | 29,01                       | 7,33                      | 46,33                     |
| 0,10                        | 21,42                                 | 7,53                        | 8,59                      | 37,54                     |
| 0,09                        | 7,70                                  | 3,26                        | 6,71                      | 17,66                     |
| 0,10                        | 9,80                                  | 10,06                       | 8,47                      | 28,33                     |
| 0,10                        | 8,36                                  | 2,81                        | 7,05                      | 18,23                     |
| 0,11                        | 24,21                                 | 13,14                       | 9,02                      | 46,37                     |
| 0,11                        | 19,63                                 | 49,10                       | 8,15                      | 76,88                     |
| 0,11                        | 8,18                                  | 4,73                        | 7,83                      | 20,74                     |
| 0,10                        | 6,44                                  | 3,85                        | 6,40                      | 16,68                     |
| 0,09                        | 8,63                                  | 26,26                       | 8,02                      | 42,91                     |
| 0,10                        | 18,47                                 | 6,57                        | 8,48                      | 33,53                     |
| 0,09                        | 6,91                                  | 3,13                        | 6,47                      | 16,50                     |
| 0,10                        | 8,60                                  | 9,93                        | 8,34                      | 26,87                     |
| 0,10                        | 8,47                                  | 2,48                        | 6,88                      | 17,83                     |
| 0,10                        | 22,76                                 | 12,10                       | 8,64                      | 43,50                     |
| 0,11                        | 17,47                                 | 43,08                       | 8,09                      | 68,64                     |
| 0,11                        | 7,75                                  | 4,42                        | 7,38                      | 19,55                     |
| 0,10                        | 6,12                                  | 3,84                        | 6,99                      | 16,95                     |
| 0,09                        | 7,47                                  | 27,72                       | 7,64                      | 42,82                     |
| 0,10                        | 19,22                                 | 6,66                        | 8,64                      | 34,52                     |
| 0,09                        | 7,24                                  | 3,20                        | 6,63                      | 17,07                     |
| 0,10                        | 6,73                                  | 9,72                        | 8,02                      | 24,47                     |
| 0,10                        | 7,42                                  | 2,57                        | 6,95                      | 16,95                     |
| 0,10                        | 22,22                                 | 12,49                       | 8,41                      | 43,12                     |
| 0,10                        | 19,85                                 | 48,39                       | 8,16                      | 76,40                     |
| 0,11                        | 8,19                                  | 4,65                        | 7,73                      | 20,58                     |
| 0,10                        | 5,91                                  | 3,61                        | 7,29                      | 16,81                     |
| 0,10                        | 6,62                                  | 26,82                       | 7,94                      | 41,37                     |
| 0,10                        | 19,06                                 | 6,89                        | 8,64                      | 34,59                     |
| 0,09                        | 7,20                                  | 2,59                        | 7,63                      | 17,41                     |
| 0,10                        | 5,58                                  | 9,80                        | 7,64                      | 23,02                     |
| 0,10                        | 7,00                                  | 2,43                        | 6,83                      | 16,26                     |
| 0,11                        | 21,99                                 | 11,05                       | 8,29                      | 41,34                     |
| 0,11                        | 22,91                                 | 50,67                       | 8,38                      | 81,96                     |
| 0,11                        | 7,54                                  | 4,31                        | 7,88                      | 19,73                     |

Tabella 10.25 – Indicatori energetici DataSet Provincie Sicilia.

| Residenziale<br>[CO <sub>2</sub> /p.ne] | Agricoltura<br>[CO <sub>2</sub> /Mln.€] | Industria<br>[CO <sub>2</sub> /Mln.€] | Terziario<br>[tCO <sub>2</sub> /Mln.€] |
|---|---|---------------------------------------|--|
| 0,54                                    | 41,08                                   | 25,44                                 | 38,92                                  |
| 0,52                                    | 49,02                                   | 185,66                                | 40,94                                  |
| 0,52                                    | 118,82                                  | 42,51                                 | 45,19                                  |
| 0,47                                    | 45,33                                   | 19,10                                 | 38,47                                  |
| 0,58                                    | 48,71                                   | 63,65                                 | 47,28                                  |
| 0,58                                    | 43,25                                   | 16,78                                 | 40,27                                  |
| 0,59                                    | 118,94                                  | 76,49                                 | 46,78                                  |
| 0,59                                    | 123,92                                  | 310,06                                | 50,90                                  |
| 0,60                                    | 39,58                                   | 28,92                                 | 43,27                                  |
| 0,51                                    | 40,36                                   | 25,18                                 | 37,11                                  |
| 0,50                                    | 53,33                                   | 179,16                                | 36,81                                  |
| 0,52                                    | 119,88                                  | 40,58                                 | 45,49                                  |
| 0,46                                    | 41,01                                   | 15,27                                 | 37,11                                  |
| 0,54                                    | 46,82                                   | 56,94                                 | 43,87                                  |
| 0,55                                    | 43,38                                   | 16,09                                 | 37,34                                  |
| 0,57                                    | 118,34                                  | 74,18                                 | 46,46                                  |
| 0,56                                    | 109,87                                  | 275,41                                | 44,91                                  |
| 0,59                                    | 39,75                                   | 25,97                                 | 40,64                                  |
| 0,49                                    | 38,79                                   | 23,98                                 | 34,00                                  |
| 0,48                                    | 52,06                                   | 151,12                                | 38,18                                  |
| 0,50                                    | 111,58                                  | 39,23                                 | 44,74                                  |
| 0,45                                    | 40,09                                   | 16,96                                 | 34,95                                  |
| 0,53                                    | 51,06                                   | 52,42                                 | 44,11                                  |
| 0,53                                    | 43,57                                   | 14,62                                 | 36,75                                  |
| 0,56                                    | 126,11                                  | 68,46                                 | 47,00                                  |
| 0,55                                    | 102,24                                  | 255,77                                | 42,48                                  |
| 0,56                                    | 42,63                                   | 24,65                                 | 40,77                                  |
| 0,46                                    | 30,92                                   | 18,48                                 | 30,71                                  |
| 0,45                                    | 41,46                                   | 126,09                                | 38,51                                  |
| 0,46                                    | 88,72                                   | 31,57                                 | 40,72                                  |
| 0,41                                    | 33,17                                   | 15,02                                 | 31,05                                  |
| 0,50                                    | 41,31                                   | 47,68                                 | 40,04                                  |
| 0,49                                    | 40,69                                   | 11,90                                 | 33,05                                  |
| 0,50                                    | 109,29                                  | 58,12                                 | 41,50                                  |
| 0,51                                    | 83,89                                   | 206,87                                | 38,87                                  |
| 0,51                                    | 37,20                                   | 21,22                                 | 35,44                                  |
| 0,44                                    | 28,19                                   | 17,67                                 | 32,18                                  |
| 0,43                                    | 34,37                                   | 127,64                                | 35,17                                  |
| 0,44                                    | 88,48                                   | 30,66                                 | 39,80                                  |
| 0,40                                    | 33,32                                   | 14,74                                 | 30,54                                  |
| 0,47                                    | 31,00                                   | 44,75                                 | 36,92                                  |
| 0,47                                    | 34,17                                   | 11,85                                 | 32,02                                  |
| 0,47                                    | 102,31                                  | 57,49                                 | 38,74                                  |
| 0,48                                    | 91,40                                   | 222,83                                | 37,59                                  |
| 0,49                                    | 37,73                                   | 21,42                                 | 35,60                                  |
| 0,45                                    | 27,03                                   | 16,50                                 | 33,31                                  |
| 0,44                                    | 30,24                                   | 122,58                                | 36,31                                  |
| 0,46                                    | 87,10                                   | 31,51                                 | 39,48                                  |
| 0,40                                    | 32,90                                   | 11,82                                 | 34,88                                  |
| 0,47                                    | 25,51                                   | 44,77                                 | 34,93                                  |
| 0,47                                    | 31,99                                   | 11,10                                 | 31,23                                  |
| 0,48                                    | 100,52                                  | 50,53                                 | 37,91                                  |
| 0,50                                    | 104,73                                  | 231,59                                | 38,32                                  |
| 0,50                                    | 34,46                                   | 19,71                                 | 36,01                                  |

Tabella 10.26 – Indicatori ambientali DataSet Provincie Sicilia.

**10.2.2 Analisi del DataSet**

Le statistiche descrittive dei dati è riportata nelle Tabelle 10.27-10.30.

|                     | T media [°C] | Popolazione | Valore_agg_Agricoltura | Valore_agg_Industria | Valore_agg_Terziario | Co <sub>2</sub> _fatt [tCO <sub>2</sub> /tep] |
|---------------------|--------------|-------------|------------------------|----------------------|----------------------|---|
| Numero vaolri       | 54           | 54          | 54                     | 54                   | 54                   | 54  |
| Valori di testo     | 0,00         | 0,00        | 0,00                   | 0,00                 | 0,00                 | 0,00  |
| Valori mancanti     | 0,00         | 0,00        | 0,00                   | 0,00                 | 0,00                 | 0,00  |
| Vaoliri unici       | 54           | 54          | 54                     | 54                   | 54                   | 54  |
| Valori nulli        | 0,00         | 0,00        | 0,00                   | 0,00                 | 0,00                 | 0,00  |
| Valore minimo       | 14,78        | 169986,50   | 139453644,00           | 316000000,00         | 1695421659,01        | 4,57  |
| Valore massimo      | 18,98        | 1253198,13  | 432473016,00           | 2636380740,00        | 17179000000,00       | 5,51  |
| Media               | 17,46        | 436190,50   | 318593869,83           | 875577682,00         | 4776457624,15        | 5,21  |
| Valore medio        | 17,05        | 559911,28   | 298664165,97           | 1256660690,29        | 6606480929,32        | 5,01  |
| Deviazione standard | 0,96         | 350674,50   | 94294864,58            | 738082045,47         | 4591501824,16        | 0,36  |

Tabella 10.27 – Statistiche descrittive valori climatici, economici ed ambientali DataSet Province Sicilia.

|                     | Residenziale | Agricoltura | Industria | Terziario |
|---------------------|--------------|-------------|-----------|-----------|
| Numero vaolri       | 54           | 54          | 54        | 54        |
| Valori di testo     | 0,00         | 0,00        | 0,00      | 0,00      |
| Valori mancanti     | 0,00         | 0,00        | 0,00      | 0,00      |
| Vaoliri unici       | 54           | 54          | 54        | 54        |
| Valori nulli        | 0,00         | 0,00        | 0,00      | 0,00      |
| Valore minimo       | 14,69        | 0,96        | 4,79      | 11,83     |
| Valore massimo      | 132,06       | 9,69        | 238,56    | 117,36    |
| Media               | 44,23        | 2,38        | 45,23     | 37,20     |
| Valore medio        | 56,43        | 4,02        | 69,03     | 51,26     |
| Deviazione standard | 35,52        | 3,04        | 63,08     | 34,92     |

Tabella 10.28 – Statistiche descrittive consumi energetici DataSet Province Sicilia.

| Anno                | Residenziale | Agricoltura | Industria | Terziario |
|---------------------|--------------|-------------|-----------|-----------|
| Numero vaolri       | 54           | 54          | 54        | 54        |
| Valori di testo     | 0,00         | 0,00        | 0,00      | 0,00      |
| Valori mancanti     | 0,00         | 0,00        | 0,00      | 0,00      |
| Vaoliri unici       | 54           | 54          | 54        | 54        |
| Valori nulli        | 0,00         | 0,00        | 0,00      | 0,00      |
| Valore minimo       | 0,09         | 5,58        | 2,43      | 6,40      |
| Valore massimo      | 0,11         | 24,21       | 56,26     | 9,24      |
| Media               | 0,10         | 8,54        | 7,57      | 7,86      |
| Valore medio        | 0,10         | 12,21       | 13,81     | 7,75      |
| Deviazione standard | 0,01         | 6,43        | 15,07     | 0,73      |

Tabella 10.29 – Statistiche indicatori energetici DataSet Province Sicilia.



| Anno                | Residenziale | Agricoltura | Industria | Terziario |
|---------------------|--------------|-------------|-----------|-----------|
| Numero vaolri       | 54           | 54          | 54        | 54        |
| Valori di testo     | 0,00         | 0,00        | 0,00      | 0,00      |
| Valori mancanti     | 0,00         | 0,00        | 0,00      | 0,00      |
| Vaoliri unici       | 54           | 54          | 54        | 54        |
| Valori nulli        | 0,00         | 0,00        | 0,00      | 0,00      |
| Valore minimo       | 0,40         | 25,51       | 11,10     | 30,54     |
| Valore massimo      | 0,60         | 126,11      | 310,06    | 50,90     |
| Media               | 0,50         | 43,48       | 39,91     | 38,49     |
| Valore medio        | 0,50         | 61,33       | 69,46     | 38,81     |
| Deviazione standard | 0,05         | 32,94       | 76,73     | 4,76      |

Tabella 10.30 – Statistiche indicatori ambientali DataSet Provincie Sicilia.

Attraverso la matrice degli scatter è possibile verificare le correlazioni tra i dati del DataSet, (Figura 10.21).

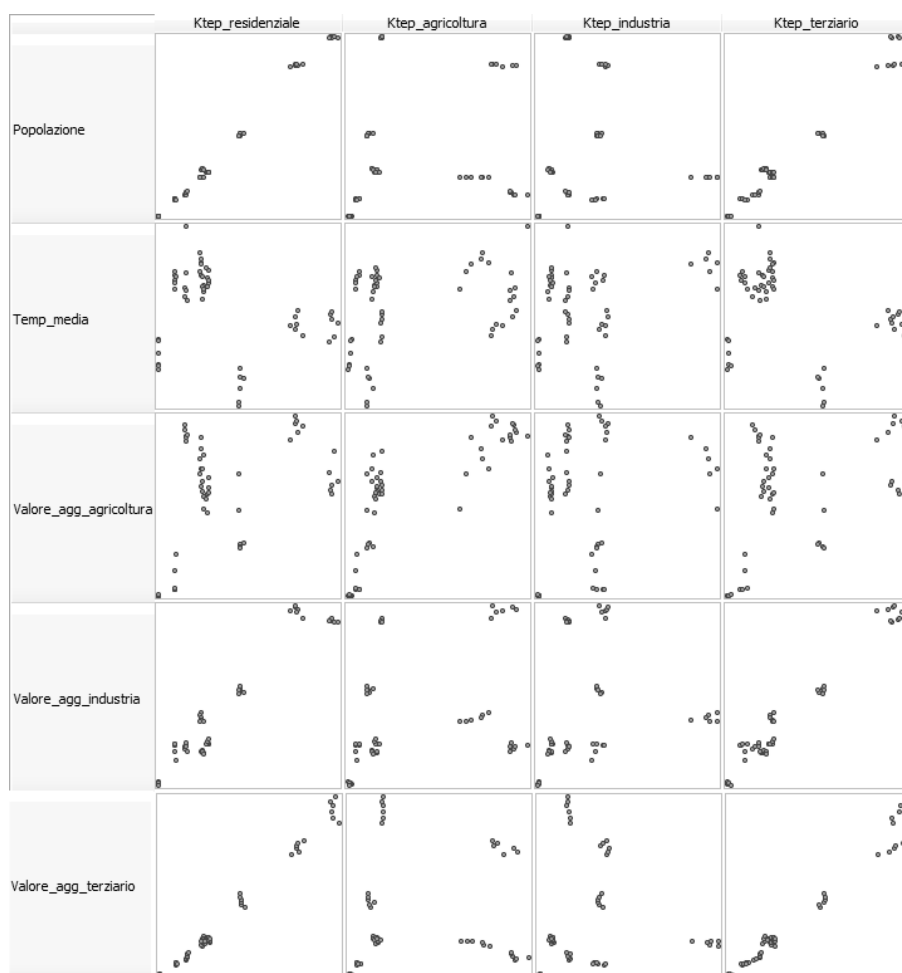


Figura 10.21 – Matrice degli scatter DataSet Provincie Sicilia.

È stato quindi verificato il valore dell'indice di correlazione per il singolo target (Figure 10.22 - 10.25). Il valore di indice positivo indica che il valore targhet cresce con il crescere del valore del fattore correlato, indici negativi indicano invece che il valore target diminuisce con il crescere del valore del fattore correlato.

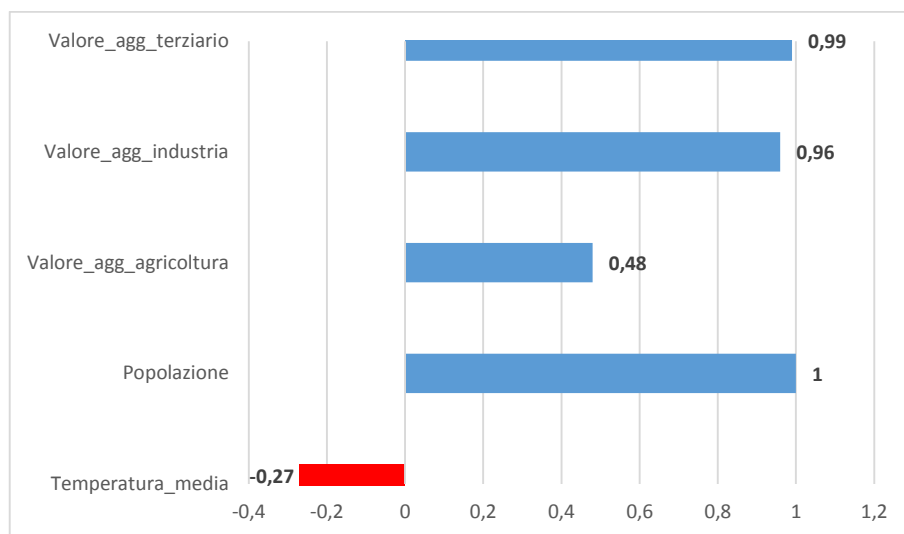


Figura 10.22 - Indici di correlazione per consumo in ktep nel residenziale.  
(■ valore negativo - ■ valori positivi)

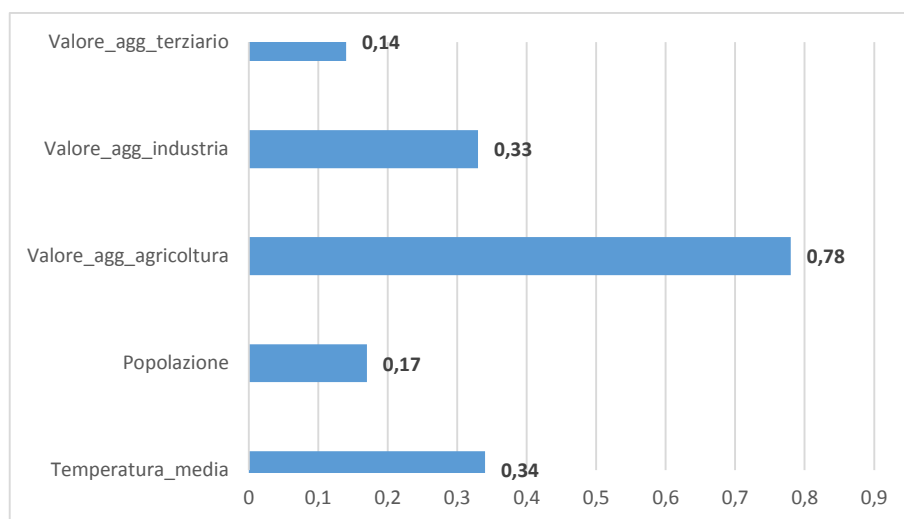


Figura 10.23 - Indici di correlazione per consumo in ktep in agricoltura.  
(■ valore negativo - ■ valori positivi)

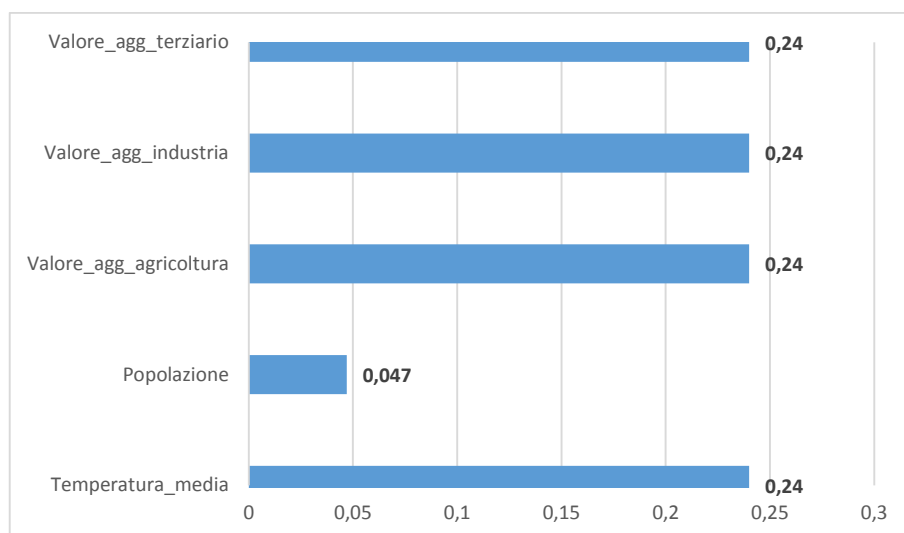


Figura 10.24 - Indici di correlazione per consumo in ktep nell'industria.

(■ valore negativo - ■ valori positivi)

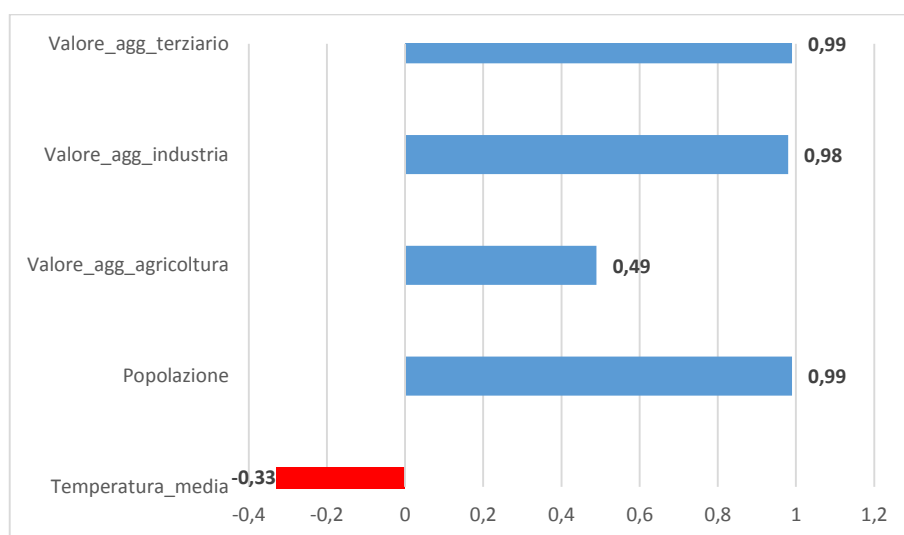


Figura 10.25 - Indici di correlazione per consumo in ktep nel terziario.

(■ valore negativo - ■ valori positivi)

### 10.2.3 Modello di previsione Neural Network “Consumi EE provinciali”

Il modello sviluppato con il software GMDH Shell basato su GMDH Neural Network ha come scopo la definizione dei consumi di energia elettrica annuale per singolo settore ponendo come input della rete neurale i dati relativi alla popolazione, al valore aggiunto per singolo settore, ed alla temperatura media annuale. Nella Figura 10.26 si riporto lo schema della rete utilizzata.

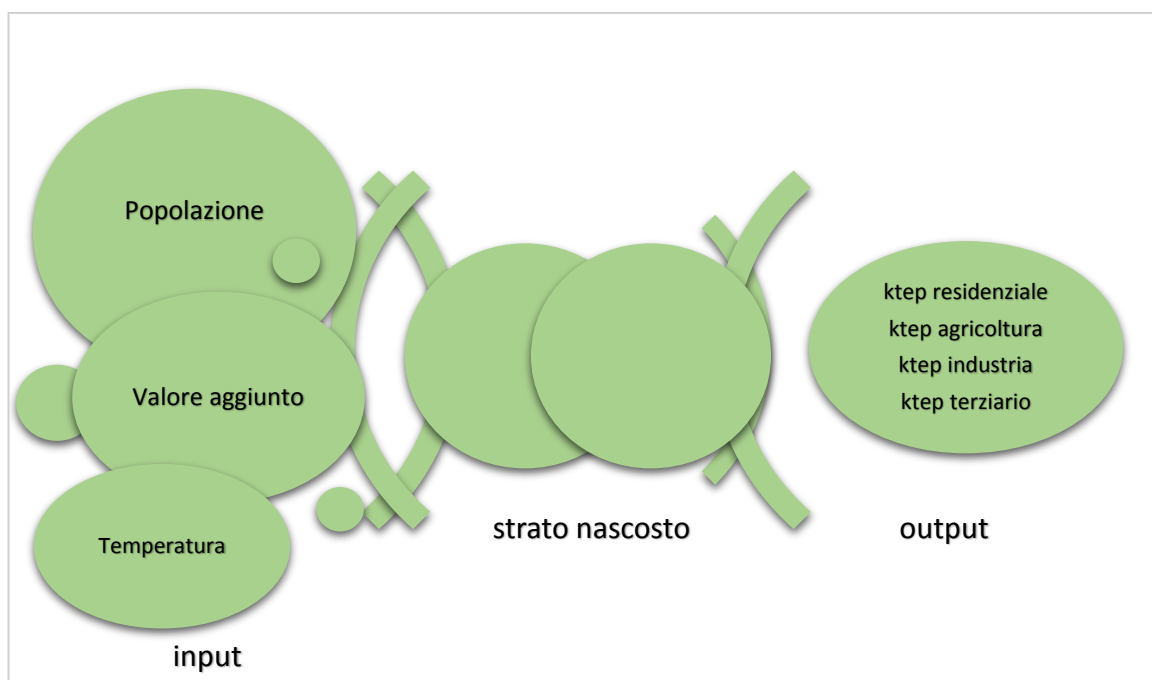


Figura 10.26 – Diagramma neurale - Previsione “Consumi EE provinciali”.

Attraverso l'iterazione di più settaggi si è scelto quello che coniugasse basso errore e velocità computazionale. Sono state quindi definite le seguenti impostazioni nel pannello Solver (Tabella 10.31).

|                       |                               |
|-----------------------|-------------------------------|
| Reorder observations  | Odd/even                      |
| Validation strategy   | k-fold validation             |
| Number of folds       | 2                             |
| Validation criterion  | RMSE                          |
| Variables ranking     | No                            |
| Core algorithm        | GMDH neural network           |
| Neuron function       | $a + x_i + x_i * x_j + x_i^2$ |
| Max. number of layers | 33                            |
| Initial layer width   | 1.000                         |
| Time series mode      | -                             |

Tab. 10.31 – Settings solver – Modello di previsione “Consumi EE provinciali”.

In particolare, il criterio di validazione è RMSE, che rappresenta la deviazione standard del campione delle differenze tra i valori previsti ed i valori osservati; la funzione neurale di attivazione del singolo neurone è  $a + x_i + x_i * x_j + x_i^2$ , con la quale si ottiene il plot del model fit e della predictions; il massimo numero di strati è pari a 33 come limite superiore per il numero di strati di rete creati dall'algoritmo; e

la larghezza dello strato iniziale è di 1.000, che definisce quanti neuroni vengono aggiunti alla serie di ingresso ad ogni nuovo livello.

Avviando il processo di training e di validation si sono quindi ottenuti i seguenti risultati per i singoli target (Tabelle 10.32 – 10.33).

| Postprocessed results                          | Model fit              | Predictions |
|--|------------------------|-------------|
| Number of observations                         | 43                     | 11          |
| Max. negative error                            | -3,59                  | -2,95       |
| Max. positive error                            | 4,83                   | 2,94        |
| Normalized Mean absolute error (nMAE)          | 1,07 %                 | 1,11 %      |
| Normalized Root mean square error (nRMSE)      | 1,44 %                 | 1,36 %      |
| Residual sum                                   | -6,39 e <sup>-14</sup> | -2,89       |
| Standard deviation of residuals                | 1,74                   | 1,58        |
| Coefficient of determination (R <sup>2</sup> ) | 0,99                   | 0,99        |
| Correlation                                    | 0,99                   | 0,99        |

Tabella 10.32 – Accuracy ktep residenziale  
Modello di previsione “Consumi EE provinciali” con Neural Network.

| Postprocessed results                          | Model fit | Predictions |
|--|-----------|-------------|
| Number of observations                         | 43        | 11          |
| Max. negative error                            | -,15      | -3,78       |
| Max. positive error                            | 1,03      | 0,92        |
| Normalized Mean absolute error (nMAE)          | 3,88 %    | 8,96 %      |
| Normalized Root mean square error (nRMSE)      | 5,08 %    | 15,26 %     |
| Residual sum                                   | -9,94e-14 | -2,39       |
| Standard deviation of residuals                | 0,41      | 1,22        |
| Coefficient of determination (R <sup>2</sup> ) | 0,98      | 0,81        |
| Correlation                                    | 0,99      | 0,90        |

Tabella 10.33 – Accuracy ktep agricoltura  
Modello di previsione “Consumi EE provinciali” con Neural Network.

| Postprocessed results                          | Model fit             | Predictions |
|--|-----------------------|-------------|
| Number of observations                         | 43                    | 11          |
| Max. negative error                            | -18,41                | -44,05      |
| Max. positive error                            | 28,41                 | 40,22       |
| Normalized Mean absolute error (nMAE)          | 2,88 %                | 6,22 %      |
| Normalized Root mean square error (nRMSE)      | 3,80 %                | 9,43 %      |
| Residual sum                                   | 1,129e <sup>-12</sup> | 62,09       |
| Standard deviation of residuals                | 8,89                  | 21,32       |
| Coefficient of determination (R <sup>2</sup> ) | 0,98                  | 0,82        |
| Correlation                                    | 0,99                  | 0,92        |

Tabella 10.34 – Accuracy ktep industria  
Modello di previsione “Consumi EE provinciali” con Neural Network.

| Postprocessed results                     | Model fit      | Predictions |
|---|----------------|-------------|
| Number of observations                    | 43             | 11          |
| Max. negative error                       | -4,78          | -4,18       |
| Max. positive error                       | 5,26           | 4,23        |
| Normalized Mean absolute error (nMAE)     | 1,18 %         | 2,00 %      |
| Normalized Root mean square error (nRMSE) | 1,65 %         | 2,40 %      |
| Residual sum                              | $24,38e^{-13}$ | 7,375       |
| Standard deviation of residuals           | 1,74           | 2,44        |
| Coefficient of determination ( $R^2$ )    | 0,99           | 0,99        |
| Correlation                               | 0,99           | 0,99        |

Tabella 10.35 – Accuracy ktep terziario  
Modello di previsione “Consumi EE provinciali” con Neural Network.

Il calcolo del MAE e del RMSE in forma normalizzata permette di superare l’errore che si genera con il MAPE ed il RMSPE per via della prossimità al valore 0 dei dati di target.

Le Figure 10.27-10.30 rappresentano il plot del model fit e della previsione. In grigio sono rappresentati i dati reali del DataSet, in blu il risultato del model fit, in rosso il risultato della previsione e la banda rosa rappresenta la banda di confidenza della previsione.

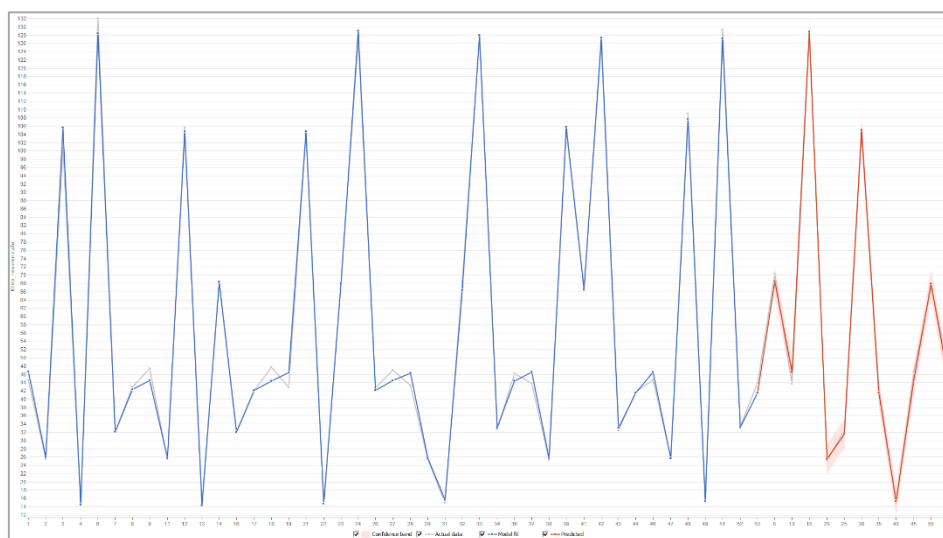


Figura 10.27 – Plot model fit e previsione ktep residenziale  
Modello di previsione “Consumi EE provinciali” con Neural Network.

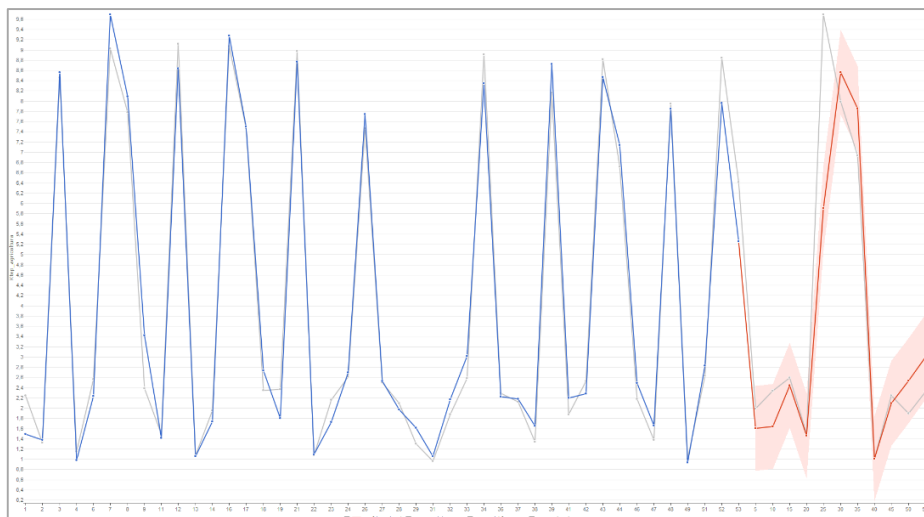


Figura 10.28 – Plot model fit e previsione ktep agricoltura  
Modello di previsione “Consumi EE provinciali” con Neural Network.

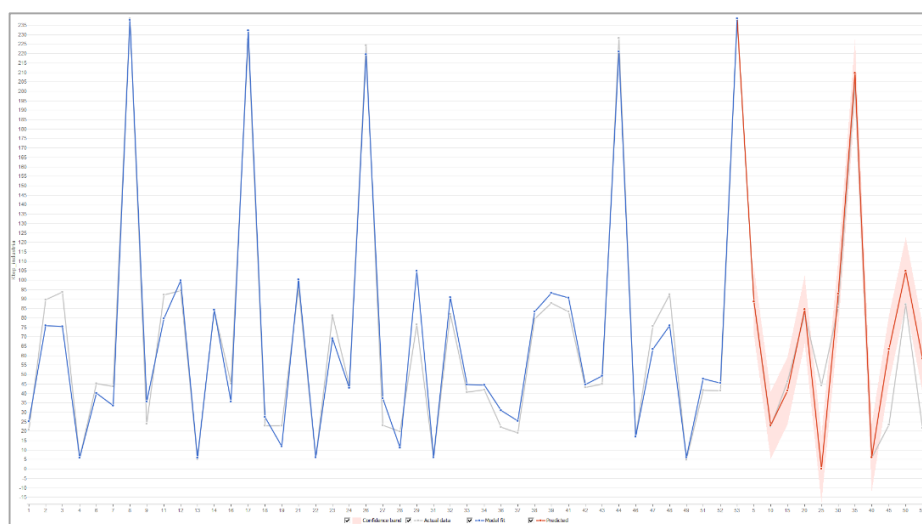


Figura 10.29 – Plot model fit e previsione ktep industria  
Modello di previsione “Consumi EE provinciali” con Neural Network.

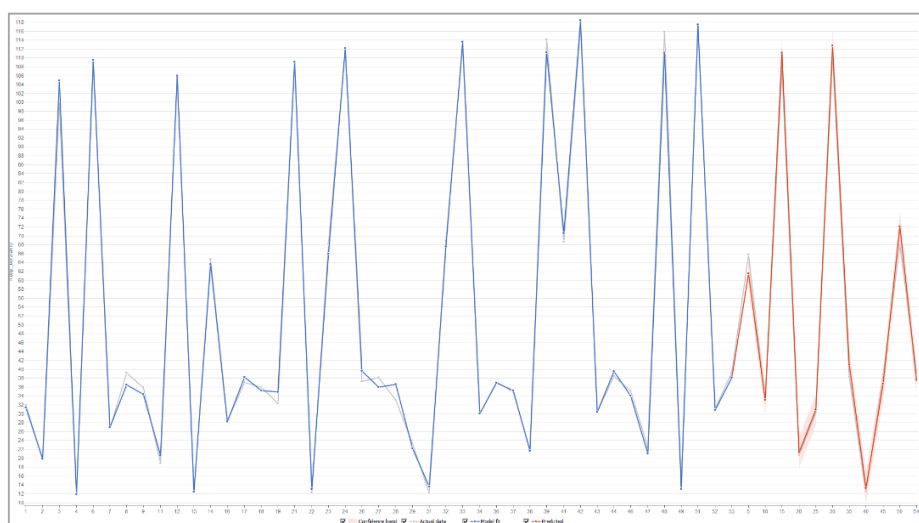


Figura 10.30 – Plot model fit e previsione ktep terziario  
Modello di previsione “Consumi EE provinciali” con Neural Network.

Di seguito si riporta le funzioni generate per la predizione.

### Consumi residenziale

$$Y1 = 0.00874723 - N79*0.865537 + N11*1.86538$$

N11 = 0.0608157 - N453\*0.211838 + N83\*1.21076  
N83 = 0.0336708 - N351\*0.926553 + N143\*1.92596  
N351 = -3.59647 + N398\*1.14675 - N398\*N466\*0.00106412  
N466 = -1.87565 + Valore\_agg\_industria\*4.6356e-08  
N398 = 1.16316 + Valore\_agg\_terziario\*1.68909e-09 + N409\*0.781302  
N453 = -1.6208 + Valore\_agg\_terziario\*9.02979e-09 - Valore\_agg\_terziario\*N468\*1.86987e-11 + N468\*0.129443  
N79 = 0.0246751 - N358\*0.941748 + N143\*1.94131  
N143 = 0.00916719 + N277\*1.50553 - N326\*0.505693  
N326 = -1.0833 - N468\*N342\*0.000503876 + N342\*1.0551  
N342 = -0.121912 + N390\*0.73008 + N455\*0.272082  
N468 = -735.704 + Temp\_medio\*43.0838 - Temp\_medio\*Valore\_agg\_agricoltura\*2.36886e-07 + Valore\_agg\_agricoltura\*4.25402e-06  
N277 = 2.12142 + N374\*0.718437 + N374\*N426\*0.000605206 + N426\*0.195823  
N426 = -6.28456 + Valore\_agg\_industria\*7.95004e-09 - Valore\_agg\_industria\*N455\*7.05451e-11 + N455\*1.05533  
N455 = -28.8919 + Temp\_medio\*1.98729 + Valore\_agg\_terziario\*7.76941e-09  
N374 = -4.04577 + Valore\_agg\_agricoltura\*1.35301e-08 - Valore\_agg\_agricoltura\*N390\*3.09096e-10 + N390\*1.10144  
N390 = -5.11217 + Popolazione\*0.000120953 - Popolazione\*Valore\_agg\_industria\*6.53204e-15  
N358 = -1.37031 + N384\*1.14625 - N384\*N459\*0.000487215 - N459\*0.083862  
N459 = 7.54496 + Valore\_agg\_industria\*N465\*1.03316e-10 + N465\*0.688376  
N465 = -51.7593 + Temp\_medio\*3.02083 - Temp\_medio\*Valore\_agg\_industria\*3.83175e-09 + Valore\_agg\_industria\*1.09909e-07  
N384 = -0.0977047 + N409\*0.698406 + N446\*0.303327  
N446 = -1.76295 + Valore\_agg\_industria\*9.15725e-09 - Valore\_agg\_industria\*Valore\_agg\_terziario\*6.8924e-19 + Valore\_agg\_terziario\*8.27204e-09  
N409 = -45.6627 + Popolazione\*0.000178126 - Popolazione\*Temp\_medio\*4.58879e-06 + Temp\_medio\*2.67769

### Consumi agricoltura

$$Y1 = -0.0273931 + N329*0.0576151 + N2*0.949054$$

N2 = 0.0173066 - N41\*0.549926 + N4\*1.54571  
N4 = -0.0352955 + N235\*0.153102 + N9\*0.855491  
N9 = 0.0240453 - N64\*0.805962 + N18\*1.80011  
N18 = -0.55509 - N258\*0.289216 - N258\*N34\*0.0420014 + N34\*1.6867  
N34 = 0.464725 + N64\*0.357089 + N64\*N122\*0.0324321 + N122\*0.325394  
N122 = -1.53813 + N443\*0.493519 - N443\*N132\*0.0606556 + N132\*1.19763  
N132 = -0.567973 - N492\*N282\*0.266015 + N282\*2.36936  
N492 = -2.69124 + Valore\_agg\_terziario\*5.4594e-10 - Valore\_agg\_terziario\*N500\*1.16336e-10 + N500\*1.60799  
N443 = 1.29888 + N458\*N480\*0.153178  
N258 = -0.376759 + N291\*1.64796 - N291\*N491\*0.119857  
N491 = -0.165642 + Valore\_agg\_industria\*8.71793e-10 + N500\*0.773596  
N291 = -1.3036 + N418\*0.644495 + N418\*N458\*0.0462574 + N458\*0.467338  
N64 = -6.12129 - N260\*2.38062 - N260\*N145\*0.428809 + N145\*7.53838  
N145 = 1.10847 + N256\*0.514634 + N256\*N265\*0.08236 - N265\*0.295061  
N265 = 0.887813 + N344\*1.46075 - N344\*N484\*0.0540231 - N484\*0.418694  
N484 = 0.85658 + N500\*N501\*0.186857  
N344 = -1.60326 + N418\*0.754867 + N418\*N453\*0.016732 + N453\*0.55957  
N256 = -0.424028 + Popolazione\*4.47854e-07 - Popolazione\*N292\*3.57208e-07 + N292\*1.27033  
N260 = -0.444131 + Valore\_agg\_industria\*1.54408e-10 - Valore\_agg\_industria\*N292\*1.49598e-10 + N292\*1.28544  
N292 = -1.51644 + N416\*0.691328 + N416\*N458\*0.0354709 + N458\*0.520155  
N416 = -0.520401 - Popolazione\*1.08079e-05 + Popolazione\*Valore\_agg\_agricoltura\*2.54846e-14 + Valore\_agg\_agricoltura\*2.01465e-08  
N235 = 1.01396 + N355\*1.71055 - N355\*N483\*0.0862137 - N483\*0.541091  
N483 = 1.35376 - N485\*0.136207 + N485\*N494\*0.179892  
N485 = 0.658826 + N496\*0.193958 + N496\*N501\*0.151384  
N355 = -0.898114 + N436\*0.534552 + N436\*N449\*0.0554721 + N449\*0.431035  
N449 = -10.5549 + Temp\_medio\*0.624694 + Temp\_medio\*N453\*0.0570246  
N453 = 1.55898 + Valore\_agg\_industria\*8.59737e-09 + Valore\_agg\_industria\*Valore\_agg\_terziario\*3.55824e-20 - Valore\_agg\_terziario\*1.31163e-09  
N436 = 0.0853937 + Valore\_agg\_agricoltura\*1.52129e-08 + Valore\_agg\_agricoltura\*N500\*3.33053e-09 - N500\*1.17752  
N41 = 0.417263 - N417\*0.274296 + N417\*N104\*0.0154144 + N104\*1.08327  
N104 = -1.40774 + N474\*0.402257 - N474\*N129\*0.0488713 + N129\*1.17403  
N129 = -0.580091 - N494\*N282\*0.275294 + N282\*2.4124  
N282 = -0.158861 + N410\*0.387079 + N410\*N447\*0.0871264 + N447\*0.246346  
N447 = -17.7738 + Temp\_medio\*1.04119 + N458\*1.00231  
N494 = -2.78353 + Popolazione\*6.94892e-06 - Popolazione\*N500\*1.43849e-06 + N500\*1.5983  
N500 = -19.6424 + Temp\_medio\*1.32368 + Valore\_agg\_terziario\*1.7636e-10  
N474 = 1.58983 + N480\*1.85692 - N496\*1.24396  
N496 = -19.7138 + Popolazione\*2.50112e-06 + Temp\_medio\*1.31404  
N480 = 1.43429 + Temp\_medio\*Valore\_agg\_industria\*1.28629e-09 - Valore\_agg\_industria\*1.96495e-08



**Consumi industria**

$$Y1 = 1.07296 - N334*0.0414148 + N2*1.02612$$

$N2 = -1.37405 - N427*N3*0.000938536 + N3*1.09827$   
 $N3 = 5.76951 - N460*0.044203 + N460*N4*0.0123963$   
 $N4 = 4.18376 - N427*0.0776864 + N5*1.01806$   
 $N5 = -14.8128 + N472*0.220046 - N472*N6*0.0061955 + N6*1.45693$   
 $N6 = 1.23367 - N169*0.246884 + N10*1.2293$   
 $N10 = 4.92088 + N12*0.842474 + N12*N17*0.000682561$   
 $N17 = 1.72964 - N213*0.354827 + N20*1.33018$   
 $N20 = -21.9753 + Popolazione*4.38793e-05 - Popolazione*N48*5.49168e-07 + N48*1.26021$   
 $N48 = 17.4217 - N423*0.342443 + N68*1.09414$   
 $N68 = 3.24532 + N138*2.43787 - N287*1.48412$   
 $N287 = -15.2069 + N294*0.963786 + N449*0.252947$   
 $N449 = -12.4585 + N459*0.858104 + N472*0.319458$   
 $N459 = -86.3303 + Popolazione*0.000202038 - Popolazione*N470*2.94858e-06 + N470*2.25445$   
 $N294 = 25.4656 - N316*0.516759 + N316*N329*0.00896679 + N329*0.164623$   
 $N316 = 13.7933 - N333*0.155045 + N333*N447*0.0153646 - N447*0.234727$   
 $N138 = 21.0197 - N219*0.816972 + N219*N460*0.022501 - N460*0.200029$   
 $N219 = 24.4278 - N321*0.681702 + N321*N327*0.0103305 + N327*0.215015$   
 $N327 = -24.2808 + Valore\_agg\_industria*1.99982e-08 - Valore\_agg\_industria*N333*3.56888e-10 + N333*1.4937$   
 $N423 = 117.777 - N454*0.734448 + N454*N469*0.0367499 - N469*2.76144$   
 $N469 = -264.276 + Temp\_media*17.0788 + Temp\_media*N474*0.0358436$   
 $N454 = -36.5983 + N470*0.803148 + N471*0.718462$   
 $N213 = 6.6665 + N261*1.04491 - N408*0.139926$   
 $N408 = 11.8328 + N433*N437*0.0125113 - N437*0.149417$   
 $N437 = 34.5571 - Valore\_agg\_agricoltura*1.94163e-07 + Valore\_agg\_agricoltura*N446*2.92594e-09 + N446*0.409578$   
 $N446 = -1214.45 + Popolazione*0.00185478 - Popolazione*Temp\_media*0.000109717 + Temp\_media*75.288$   
 $N433 = -99.314 + N465*1.4837 - N465*N470*0.0131649 + N470*1.8712$   
 $N261 = 26.7493 + N322*0.243565 + N322*N330*0.00791496 - N330*0.441394$   
 $N330 = 62.4592 - N356*0.314482 + N356*N470*0.0194657 - N470*1.01711$   
 $N470 = -264.276 + Temp\_media*18.0353 + Temp\_media*Valore\_agg\_agricoltura*5.22206e-09$   
 $N322 = -27.8383 + Valore\_agg\_terziario*3.95288e-09 - Valore\_agg\_terziario*N333*8.03253e-11 + N333*1.55997$   
 $N12 = 1.76642 + N467*N22*0.00849861 + N22*0.339068$   
 $N22 = 3.91879 - N231*1.38538 + N54*2.32953$   
 $N54 = -2.556 - N442*0.149705 - N442*N99*0.00936454 + N99*1.97353$   
 $N99 = 5.84975 - N460*0.0374902 + N460*N141*0.0123721$   
 $N141 = 17.5572 + N323*0.177207 + N323*N296*0.00510593$   
 $N296 = 31.6873 - N312*0.561172 + N312*N463*0.0196864 - N463*0.357028$   
 $N463 = -29.7458 + Popolazione*5.03978e-05 - Popolazione*N465*8.83559e-07 + N465*1.52733$   
 $N312 = 391.57 - Temp\_media*23.2436 + Temp\_media*N333*0.435418 - N333*6.39413$

**Consumi terziario**

$$Y1 = 1.16176 - Valore\_agg\_agricoltura*5.05261e-09 + N5*1.00675$$

$N5 = -0.576671 - N343*0.574649 - N343*N96*0.000213974 + N96*1.60219$   
 $N96 = 0.0896663 + N235*2.47105 - N359*1.4728$   
 $N359 = -6.22941 + N476*0.604915 - N476*N481*0.00170538 + N481*0.643842$   
 $N481 = 7.80169e-14 + N480*1$   
 $N476 = 7.93842 - Valore\_agg\_industria*6.9516e-09 + Valore\_agg\_industria*N484*1.15967e-10 + N484*0.810112$   
 $N235 = 4.54654 + N466*0.439812 + N466*N478*0.00193202 + N478*0.326771$   
 $N478 = -4.0777 + Valore\_agg\_agricoltura*1.73837e-08 + N484*0.978339$   
 $N466 = -11.3162 + Valore\_agg\_terziario*5.88838e-09 - Valore\_agg\_terziario*N480*2.93054e-11 + N480*0.75$   
 $N480 = 1.78547 + Valore\_agg\_terziario*7.48238e-09$   
 $N343 = -2.29202 + N388*0.806374 - N388*N471*0.000680856 + N471*0.289783$   
 $N471 = -1.1302 + Valore\_agg\_terziario*6.64624e-09 + N491*0.16474$   
 $N491 = -696.12 + Temp\_media*40.3302 - Temp\_media*Valore\_agg\_agricoltura*2.35574e-07 + Valore\_agg\_agricoltura*4.23913e-06$   
 $N388 = -0.853895 + N468*0.582121 + N484*0.434538$   
 $N484 = -5.23863 - Temp\_media*Valore\_agg\_industria*3.11104e-09 + Valore\_agg\_industria*9.76219e-08$   
 $N468 = -1.16655 + Popolazione*8.07373e-05 + Popolazione*Valore\_agg\_agricoltura*4.91886e-14 - Valore\_agg\_agricoltura*6.19369e-09$

**10.3 Scala comunale**

Il caso studio relativo ai comuni è volto alla definizione di un modello per la predizione dei consumi di energia elettrica e gas naturale. Tale modello, di carattere regressivo con Neural Network, è stato costruito partendo da dati annuali relativi a 385 comuni della regione Sicilia. Non sono stati presi in considerazione i dati relativi alle isole minori.

Non è stato invece possibile sviluppare un modello di serie temporali, poiché il DataSet non ha una sequenza temporale definita.

I dati di input della rete neurale sono relativi al posizionamento geografico del comune, alle caratteristiche climatiche del luogo, al numero della popolazione, ed alla penetrazione delle infrastrutture di distribuzione del gas. Sicilia.

### 10.3.1 Il DataSet

Tramite i dati provenienti dal Enel Spa [46], TERNA Spa [38], GSE Spa [39-42], ISTAT [43] è stato possibile la costruzione del database regionale, di cui un campione per 20 comuni si riporta nelle Tabelle 10.36-10.37.

| Popolazione | Superficie (Kmq) | Altitudine_ (km) | T_CUM [°C] | GG   | Radiazione (kWh/m <sup>2</sup> ) | Lunghezza_ rete_ (km) | Comune montano |
|-------------|------------------|------------------|------------|------|----------------------------------|-----------------------|----------------|
| 45237       | 130,77           | 0,258            | 91639      | 1140 | 1214,13                          | 90                    | 0              |
| 3161        | 72,72            | 0,249            | 93573      | 1086 | 1213,05                          | 0                     | 0,5            |
| 7292        | 154,79           | 0,338            | 91327      | 1230 | 1216,65                          | 30                    | 0              |
| 10763       | 65,77            | 0,11             | 97082      | 1038 | 1226,7                           | 42                    | 0              |
| 14863       | 127,15           | 0,026            | 101983     | 820  | 1003,82                          | 0                     | 0,5            |
| 30394       | 207,08           | 0,187            | 95138      | 1055 | 1224,91                          | 90                    | 0              |
| 5150        | 69,6             | 0,186            | 95163      | 1015 | 1235,2                           | 0                     | 1              |
| 28763       | 47,29            | 0,751            | 80902      | 1648 | 1255,38                          | 60                    | 0,5            |
| 4468        | 45,02            | 0,233            | 94128      | 1263 | 1220,24                          | 25                    | 0              |
| 82337       | 241,73           | 0,012            | 99555      | 816  | 1220,6                           | 68                    | 0              |
| 51369       | 275,68           | 0,008            | 99656      | 927  | 1225,99                          | 91                    | 0              |
| 11259       | 58,36            | 0,036            | 98949      | 848  | 1214,13                          | 29                    | 0              |
| 11373       | 82,43            | 0,414            | 89408      | 1236 | 1223,47                          | 43                    | 0              |
| 1649        | 37,53            | 0,189            | 93380      | 936  | 1222,03                          | 12                    | 0              |
| 1771        | 41,68            | 0,171            | 93835      | 1297 | 1222,03                          | 12                    | 0              |
| 11274       | 181,72           | 0,446            | 88601      | 1203 | 1219,88                          | 50                    | 0              |
| 5234        | 63,81            | 0,41             | 89509      | 1231 | 1221,67                          | 25                    | 0              |
| 4108        | 59,68            | 0,006            | 97009      | 813  | 1125,67                          | 0                     | 1              |
| 70648       | 272              | 0,003            | 99782      | 810  | 1265,47                          | 90                    | 0              |
| 11802       | 52,92            | 0,24             | 93800      | 1076 | 1255,38                          | 32                    | 0,5            |

Tabella 10.36 – Dati input DataSet Comuni Sicilia.

| Gas_distribuito<br>[tep] | Energia elettrica<br>[tep] |
|--------------------------|----------------------------|
| 5445                     | 4427                       |
| 0                        | 399                        |
| 733                      | 728                        |
| 2244                     | 1245                       |
| 0                        | 1764                       |
| 4455                     | 3083                       |
| 0                        | 642                        |
| 2626                     | 3056                       |
| 520                      | 435                        |
| 5705                     | 8557                       |
| 5828                     | 5334                       |
| 537                      | 1154                       |
| 1392                     | 1088                       |
| 240                      | 154                        |
| 217                      | 159                        |
| 1289                     | 1113                       |
| 625                      | 504                        |
| 0                        | 793                        |
| 7582                     | 7161                       |
| 838                      | 1307                       |

Tabella 10.37 – Dati output DataSet Comuni Sicilia.

### 10.3.2 Analisi del DataSet

Le statistiche descrittive dei dati è riportata nelle Tabelle 10.38-10.39

|                      | Popolazione | Superficie<br>(Km <sup>2</sup> ) | Altitudine<br>[km] | T_CUM<br>[°C] | GG      | Radiazione<br>(kWh/m <sup>2</sup> ) | Lunghezza_<br>rete_(km) | Comune<br>montano |
|----------------------|-------------|----------------------------------|--------------------|---------------|---------|-------------------------------------|-------------------------|-------------------|
| Numero vaolri        | 3080        | 3080                             | 3080               | 3080          | 3080    | 3080                                | 3079                    | 3080              |
| Valori di testo      | 0           | 0                                | 0                  | 0             | 0       | 0                                   | 0                       | 0                 |
| Valori mancanti      | 0           | 0                                | 0                  | 0             | 0       | 0                                   | 1                       | 0                 |
| Vaoliri unici        | 2726        | 376                              | 278                | 356           | 324     | 218                                 | 167                     | 3                 |
| Valori nulli         | 0           | 0                                | 0                  | 0             | 0       | 0                                   | 550                     | 1632              |
| Valore più frequente | 927         | 6,6                              | 0,35               | 100760        | 1396    | 1212,33                             | 0                       | 0                 |
| Valore minimo        | 11,28       | 1,14                             | 0,001              | 68547         | 579     | 953,35                              | 0                       | 0                 |
| Valore massimo       | 678492      | 551,12                           | 1,275              | 110655        | 3309    | 1294,47                             | 891                     | 1                 |
| Media                | 4762,5      | 37,27                            | 0,4                | 94138         | 1251    | 1221,32                             | 23                      | 0                 |
| Valore medio         | 13015,20    | 66,17                            | 0,40               | 93283,99      | 1268,95 | 1220,72                             | 31,71                   | 0,36              |
| Deviazione standard  | 40640,49    | 80,20                            | 0,28               | 7220,05       | 427,94  | 33,69                               | 54,74                   | 0,42              |

Tabella 10.38 – Statistiche descrittive valori input DataSet Comuni Sicilia.

|                      | Gas_distribuito<br>[tep] | Energia elettrica<br>[tep] |
|----------------------|--------------------------|----------------------------|
| Numero vaolri        | 3080                     | 3080                       |
| Valori di testo      | 0                        | 0                          |
| Valori mancanti      | 0                        | 0                          |
| Vaoliri unici        | 1466                     | 1487                       |
| Valori nulli         | 614                      | 0                          |
| Valore più frequente | 0                        | 108                        |
| Valore minimo        | 0                        | 13                         |
| Valore massimo       | 75875                    | 67458                      |
| Media                | 368                      | 453                        |
| Valore medio         | 1299,32                  | 1244,93                    |
| Deviazione standard  | 4292,65                  | 3952,59                    |

Tabella 10.39 – Statistiche descrittive output DataSet Comuni Sicilia.

Attraverso la matrice degli scatter è possibile verificare le correlazioni tra le differenti variabili del DataSet, (Figura 10.31).

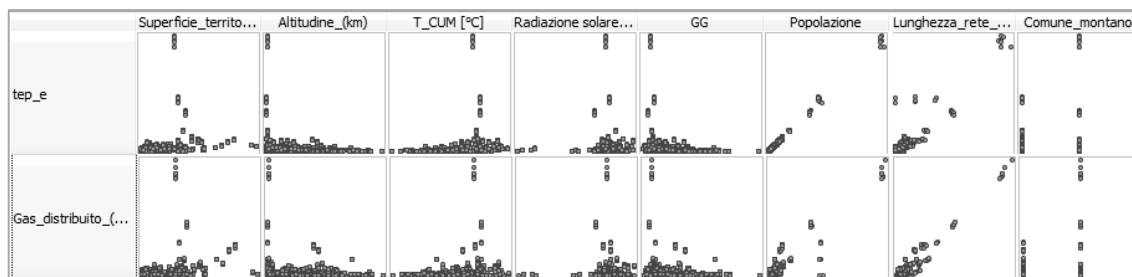


Figura 10.31 – Matrice degli scatter DataSet Sicilia.

È stato quindi verificato il valore dell'indice di correlazione per il singolo target (Figure 10.32 - 10.33). Il valore di indice positivo indica che il valore targhet cresce con il crescere del valore del fattore correlato, indici negativi indicano invece che il valore target diminuisce con il crescere del valore del fattore correlato.

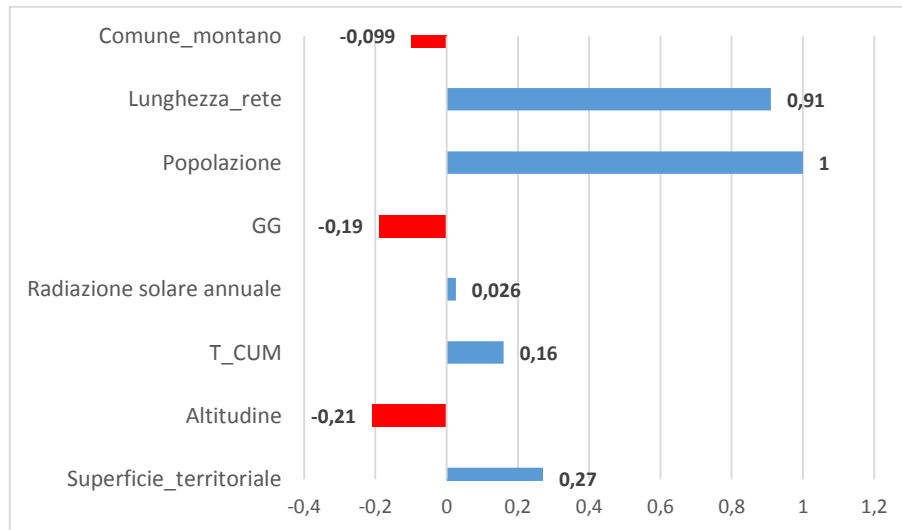


Figura 10.32 - Indici di correlazione per consumo in tep di energia elettrica.

(■ valore negativo - ■ valori positivi)

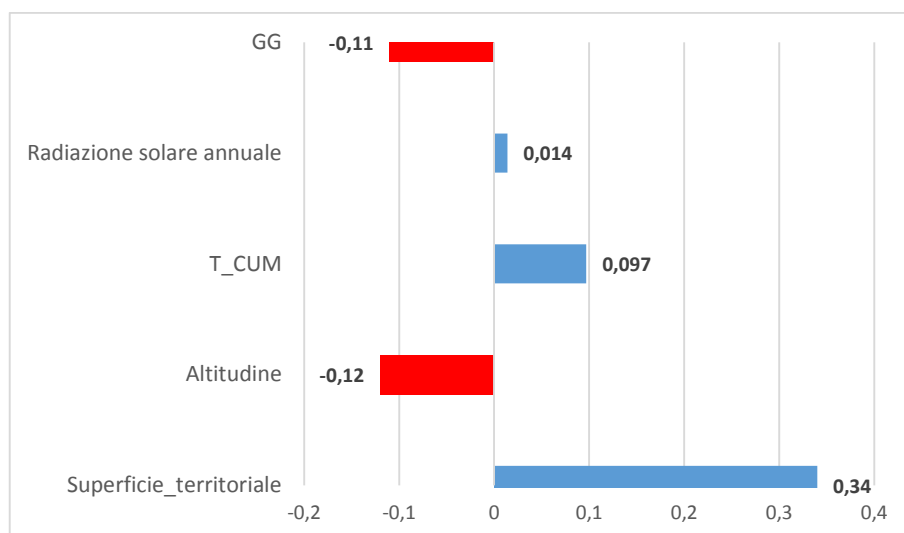


Figura 10.33 - Indici di correlazione per consumo in tep di gas.

(■ valore negativo - ■ valori positivi)

Sempre per i target è stata elaborata graficamente la correlazione con il numero della popolazione e la penetrazione della rete di distribuzione del gas (Figure 11.34-10.35).

Dalle analisi realizzate si verifica che i valori di consumo di energia elettrica e di gas naturale, superato un certo valore sia di popolazione che di km di rete di distribuzione di gas, crescono proporzionalmente.

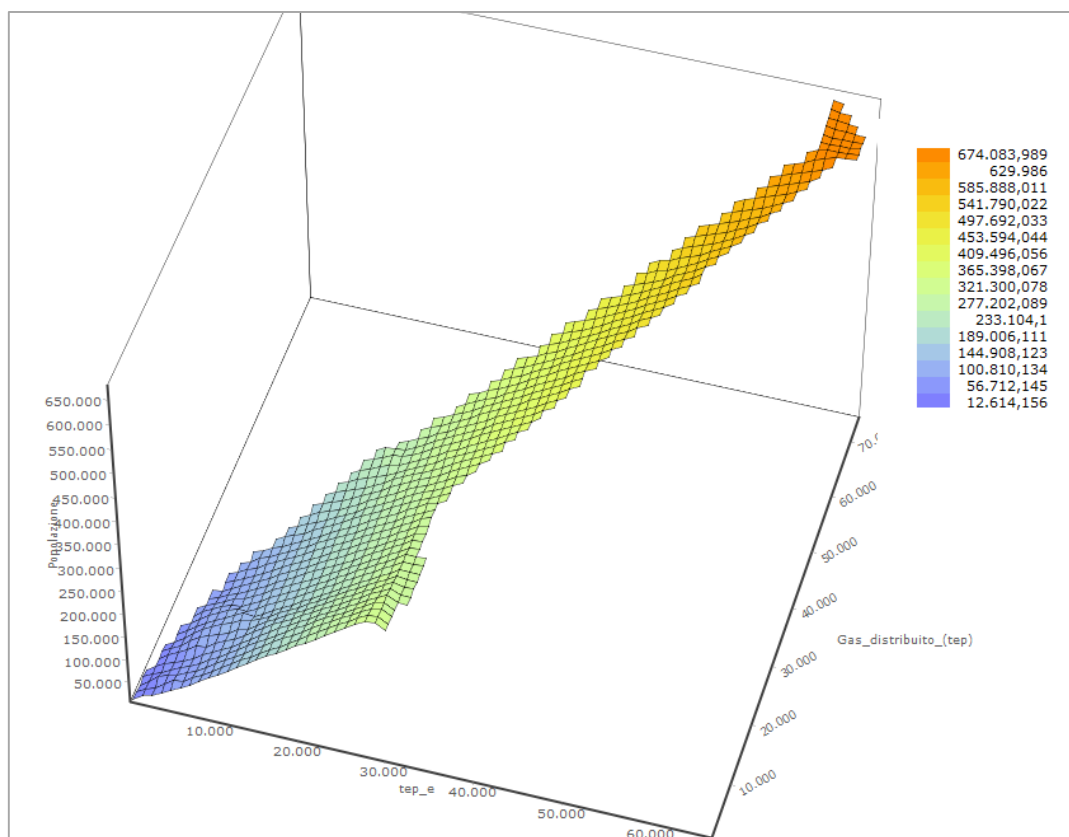


Figura 10.34 - Correlazione tra tep energia elettrica, tep gas naturale e popolazione.

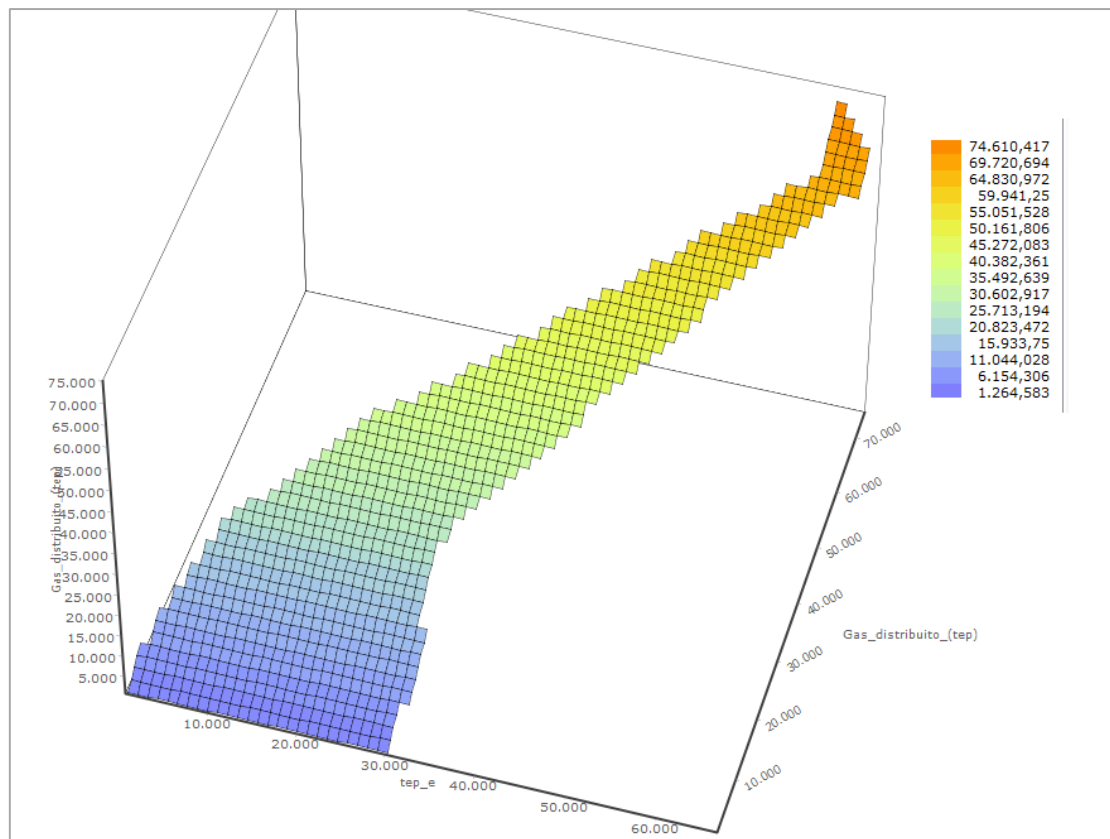


Figura 10.35 - Correlazione tra tep energia elettrica, tep gas naturale e km di rete gas.

### 10.3.3 Modello di previsione Neural Network “Consumi EE e gas comunali”

Il modello sviluppato con il software GMDH Shell basato sulle Neural Network GMDH ha come scopo la definizione dei consumi di energia elettrica e  $m^3$  di gas annuali per comune come input della rete neurale i dati relativi al posizionamento geografico del comune, alle caratteristiche climatiche del luogo, al numero della popolazione, ed alla penetrazione delle infrastrutture di distribuzione del gas. Sicilia. Nella Figura 10.36 si riporto lo schema della rete utilizzata.

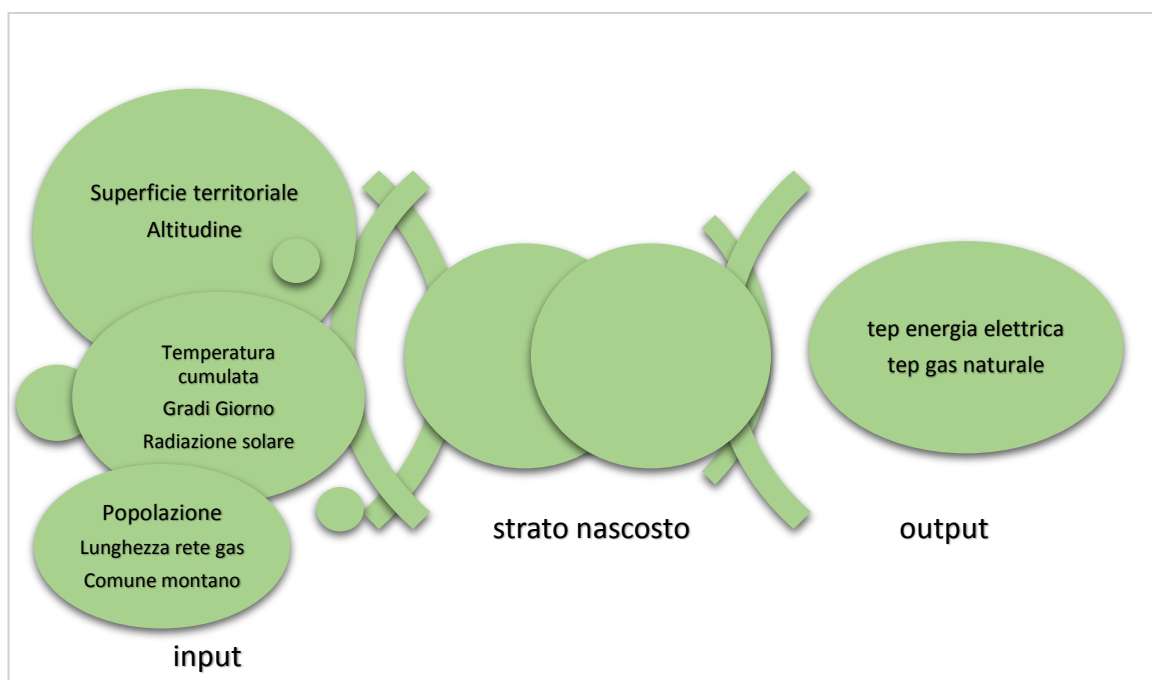


Figura 10.36 – Diagramma neurale - Previsione “Consumi EE e gas comunali”

Attraverso l'iterazione di più settaggi si è scelto quello che coniugasse basso errore e velocità computazionale. Sono state quindi definite le seguenti impostazioni nel pannello Solver (Tabella 10.40).

|                       |                       |
|-----------------------|-----------------------|
| Reorder observations  | Odd/even              |
| Validation strategy   | k-fold validation     |
| Number of folds       | 2                     |
| Validation criterion  | RMSE                  |
| Variables ranking     | No                    |
| Core algorithm        | GMDH neural network   |
| Neuron function       | $a+x_i+x_i*x_j+x_i^2$ |
| Max. number of layers | 10                    |
| Initial layer width   | 50                    |
| Time series mode      | -                     |

Tabella 10.40 – Settings solver – Modello di previsione “Consumi EE e gas comunali”.

In particolare da rilevare: il criterio di validazione RMSE, che rappresenta la deviazione standard del campione delle differenze tra i valori previsti ed i valori osservati, il core algorithm (GMDH neural network) che rappresenta l'algoritmo di apprendimento statistico, la funzione neurale di attivazione del singolo neurone ( $a+x_i+x_i*x_j+x_i^2$ ) con la quale si ottiene il plot del model fit e della predictions, il massimo numero di strati (10) come limite superiore per il numero di strati di rete creati

dall'algoritmo, e la larghezza dello strato iniziale (50) che definisce quanti neuroni vengono aggiunti alla serie di ingresso ad ogni nuovo livello.

Avviando il processo di training e di validation si sono quindi ottenuti i seguenti risultati per i singoli target (Tabelle 10.41 – 10.42).

| Postprocessed results                          | Model fit            | Predictions |
|--|----------------------|-------------|
| Number of observations                         | 2464                 | 616         |
| Max. negative error                            | -3311,4              | -1218,76    |
| Max. positive error                            | 5036,43              | 2830,39     |
| Normalized Mean absolute error (nMAE)          | 0,13 %               | 0,13 %      |
| Normalized Root mean square error (nRMSE)      | 0,34 %               | 0,30 %      |
| Residual sum                                   | 4,42 e <sup>-9</sup> | -2629,27    |
| Standard deviation of residuals                | 230,99               | 202,56      |
| Coefficient of determination (R <sup>2</sup> ) | 0,99                 | 0,99        |
| Correlation                                    | 0,99                 | 0,99        |

Tabella 10.41 – Accuracy tep energia elettrica  
Modello di previsione “Consumi EE e gas comunali” con Neural Network.

| Postprocessed results                          | Model fit | Predictions |
|--|-----------|-------------|
| Number of observations                         | 2464      | 616         |
| Max. negative error                            | -8579,57  | -8520,57    |
| Max. positive error                            | 8137,57   | 7916,19     |
| Normalized Mean absolute error (nMAE)          | 0,80 %    | 0,78 %      |
| Normalized Root mean square error (nRMSE)      | 1,50 %    | 1,46 %      |
| Residual sum                                   | 83092,9   | 40263,5     |
| Standard deviation of residuals                | 1139,63   | 1107,22     |
| Coefficient of determination (R <sup>2</sup> ) | 0,93      | 0,90        |
| Correlation                                    | 0,96      | 0,95        |

Tabella 10.42 – Accuracy tep gas naturale  
Modello di previsione “Consumi EE e gas comunali” con Neural Network.

Il calcolo del MAE e del RMSE in forma normalizzata permette di superare l'errore che si genera con il MAPE ed il RMSPE per via della prossimità al valore 0 dei dati di target.



Le Figure 10.37-10.38 rappresentano il plot del model fit e della predizione. In grigio sono rappresentati i dati reali del DataSet, in blu il risultato del model fit, in rosso il risultato della predizione e la banda rosa rappresenta la banda di confidenza della predizione.

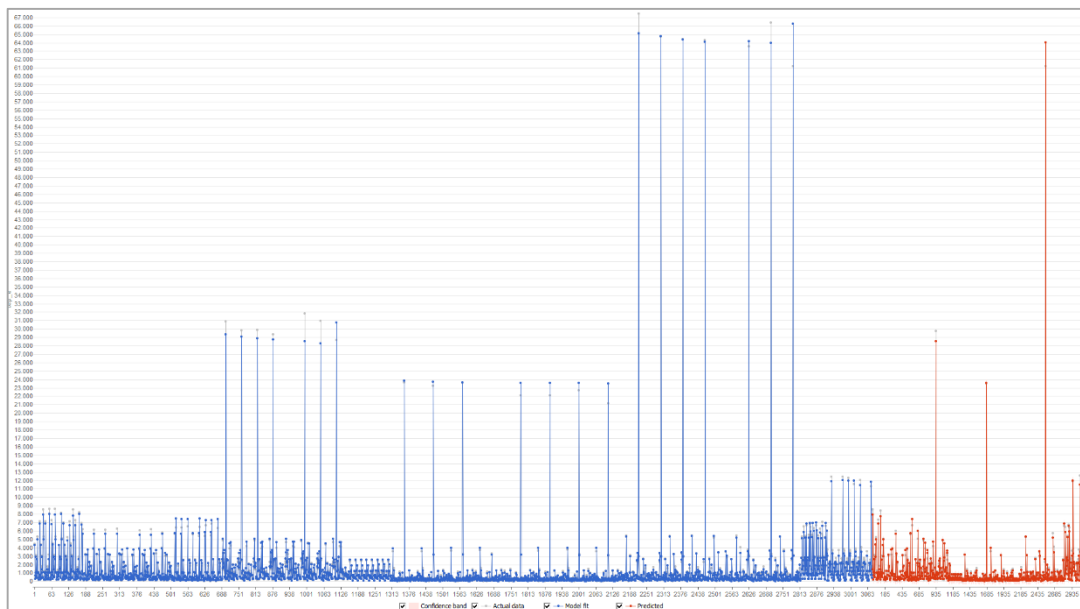


Figura 10.37 – Plot model fit e previsione tep energia elettrica  
Modello di previsione “Consumi EE e gas comunali” con Neural Network.

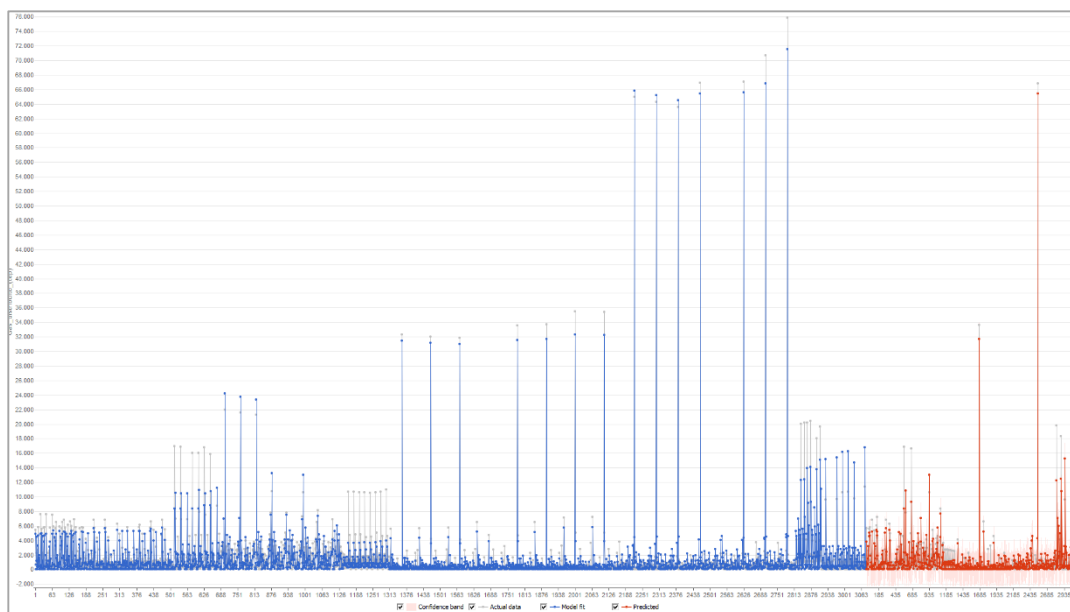


Figura 10.38 – Plot model fit e previsione tep gas naturale  
Modello di previsione “Consumi EE e gas comunali” con Neural Network.

Nelle figure 10.39 – 10.44 sono mostrati rispettivamente il “residuals” del model fit e della predictions, l’autocorrelazione e la frequenza del valore residuo.

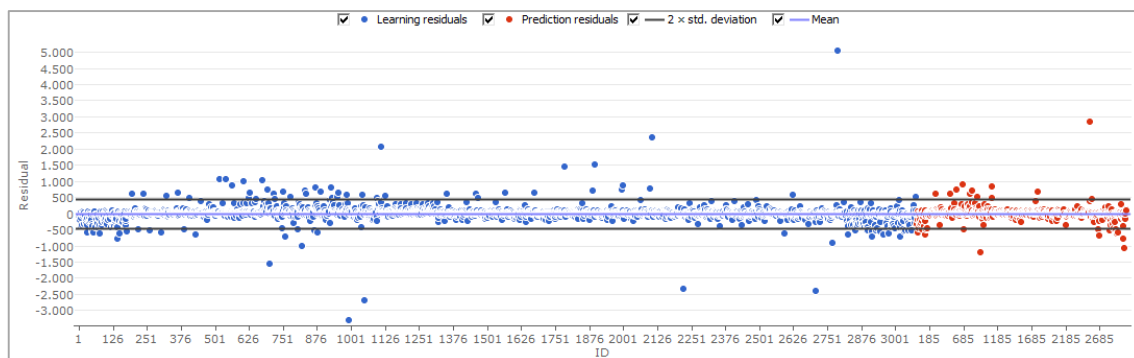


Figura 10.39 – Plot residui previsione tep energia elettrica  
“Consumi EE e gas comunali” con Neural Network

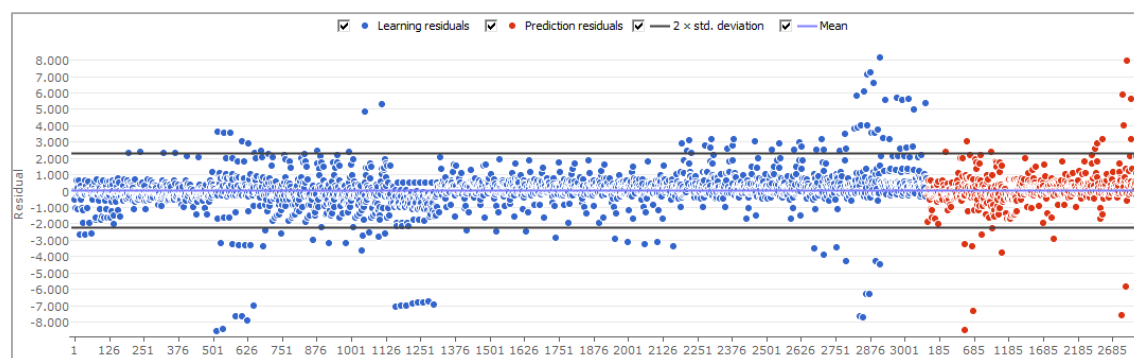


Figura 10.40 – Plot residui previsione tep gas naturale  
“Consumi EE e gas comunali” con Neural Network.

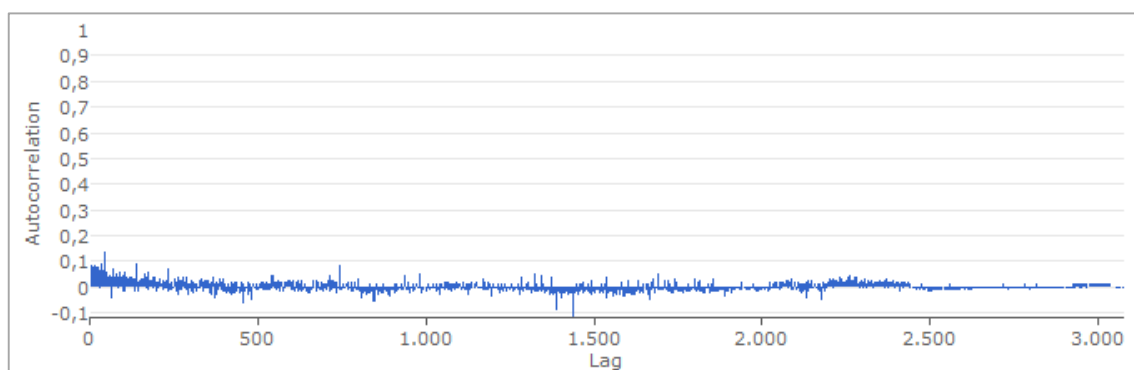


Figura 10.41 – Autocorrelazione previsione tep energia elettrica  
“Consumi EE e gas comunali” con Neural Network.

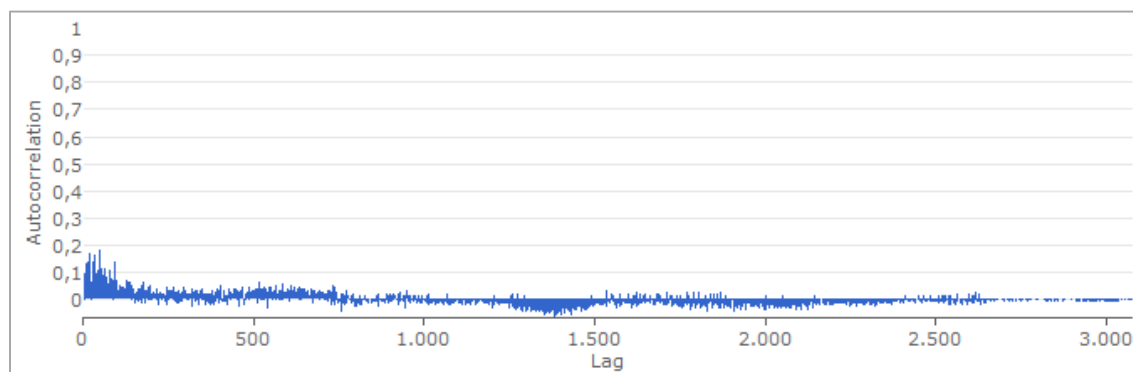


Figura 10.42 – Autocorrelazione previsione tep gas naturale  
“Consumi EE e gas comunali” con Neural Network.

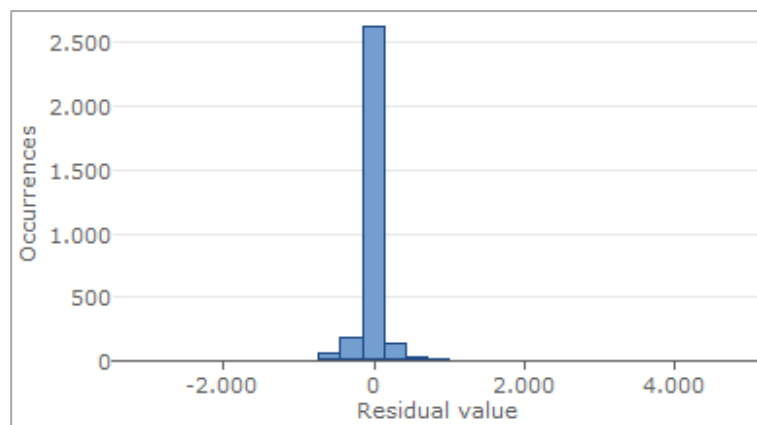


Figura 10.43 – Frequenza residui previsione tep energia elettrica  
“Consumi EE e gas comunali” con Neural Network.

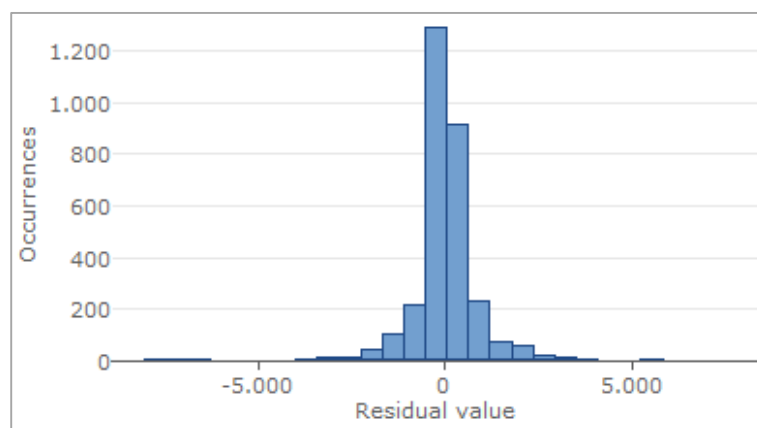


Figura 10.44 – Frequenza residui previsione tep gas naturale  
“Consumi EE e gas comunali” con Neural Network.

Di seguito si riporta le funzioni generate per la predizione.

#### **Consumi di energia elettrica**

$$Y1 = 12.1279 + N9 * 1.01421 - N25 * 0.0237402$$

$$N25 = -6174.67 + \text{"Popolazione, cubert"} * 363.528 + \text{"Comune\_montano, cubert"} * 1286.42$$

$$N9 = 101.464 + \text{"Popolazione"} * 0.0971303 - \text{"Altitudine\_ (km), cubert"} * 181.142$$

#### **Consumi di gas naturale**

$$Y1 = 114.343 - N67 * 0.104056 + N9 * 1.01799$$

$$N9 = -23.0468 + N24 * 0.821721 + N42 * 0.195627$$

$$N42 = -156.432 + N44 * 0.955796 + N66 * 0.161952$$

$$N66 = 5158.32 - \text{"Radiazione solare annuale (kWhICUM2/m2)"} * 6.54641 + \text{"Lunghezza\_rete\_ (km), cubert"} * 1622.18$$

$$N44 = -257.216 + \text{"Superficie\_territoriale(Kmq)"} * 4.95152 + \text{"Popolazione"} * 0.094594$$

$$N24 = -359.895 + \text{"Altitudine\_ (km)"} * 888.003 + N33 * 1.01005$$

$$N33 = 314.47 + \text{"Lunghezza\_rete\_ (km)"} * 86.182 - \text{"Lunghezza\_rete\_ (km), cubert"} * 680.498$$

$$N67 = 367.389 + \text{"Superficie\_territoriale(Kmq), cubert"} * 1168.28 - \text{"Altitudine\_ (km), cubert"} * 4780.72$$

## 10.4 Conclusioni

L'analisi dei tre sotto casi studio ha confermato la migliore resa dei sistemi predittivi basati su Neural Network quando il DataSet è ben strutturato e rappresentativo di un campione di dati numeroso.

È infatti risultato più vantaggiosa l'utilizzo di un DataSet contenente dati provinciali, rispetto a quello contenente dati regionali. Infatti, grazie ai dati disaggregati, nel caso della Sicilia in nove sotto-insiemi (province), sono stati utilizzati come input molti più dati per l'allenamento della rete e di conseguenza sono migliorati i parametri di bontà del modello predittivo. Pertanto, risulta preferibile nel caso anche di predizione degli indici energetici regionali, procedere attraverso la predizione dei parametri provinciali, e successivamente per aggregazione addivenire al dato Regionale.

Per quanto riguarda invece i dati Comunali, i risultati ottenuti hanno evidenziato come la buona organizzazione di un DataSet garantisca al contempo buoni parametri di bontà della rete (nMAE e nRMSE).

Infine, è evidente come i modelli su serie storiche siano consigliabili esclusivamente per analisi relative ad un unico soggetto (Regione), e non per più soggetti come nel caso di province e comuni. Questo perché l'organizzazione del DataSet genera una perdita della sequenza temporale che non rende applicabile alcun modello di previsione basato esclusivamente sulla periodicità di eventi e stagionalità dell'output.

A livello regionale risulta nel modello predittivo Neural Network che la scelta di dati di input quali popolazione, valore aggiunto e temperatura:

- rendono il modello molto affidabile per la predizione dei consumi nel settore industriale;
- buono per la predizione per i settori residenziale, terziario, servizi non vendibili ed attività produttive;
- non ottimale per la quantificazione dei consumi nel settore agricolo. Pertanto, per tale settore risulta preferibile la creazione di un modello specifico che utilizzi altre variabili come input del modello predittivo.

Analizzando a livello regionale i risultati dei modelli di previsione basati sulle serie storiche è evidente il discostarsi per eccesso di consumo in tutti i settori nei sei anni 2008-2013 della previsione rispetto al dato registrato; tuttavia il trend di crescita per i settori residenziale, terziario ed agricoltura risulta confermato anche dai dati reali, così come anche i trend di depressione o stagnazione per l'industria e le attività produttive. Risulta invece non conforme la predizione dei servizi non vendibili rispetto ai dati reali; tale risultato è giustificato dalla contrazione dei servizi resi dalla pubblica amministrazione nel triennio al fine di diminuire la spesa pubblica (Spending Review). Nel complesso a parte l'industria, che registra una graduale stagnazione, i modelli di previsione su serie storiche indicano per la regione Sicilia un costante aumento dei consumi in tutti gli altri settori.

A livello provinciale, gli indici utilizzati quali input al modello di previsione basato su Neural Network, confermano la loro bontà – come avvenuto anche a livello regionale – nei settori residenziale, industria e terziario. L'agricoltura conferma la necessità di utilizzare come input al modello variabili che meglio possono descrivere le dinamiche energetiche del settore.

A livello comunale, le previsioni del consumo di energia elettrica con modello Neural Network forniscono previsioni più accurate rispetto alla previsione del consumo di gas metano. Pur tuttavia, gli indici utilizzati come input al modello (superficie territoriale, altitudine, temperatura cumulata, gradi giorno, radiazione solare, popolazione, lunghezza rete gas, indice di comune montano) risultano corretti al fine di fornire buoni livelli di affidabilità del modello.

## CAPITOLO 11 *Caso studio*

### *Produzione di energia elettrica da FER*

#### **11.1 Il contesto della regione Sicilia**

Il bilancio della precedente programmazione (PEARS 2009) si chiude in positivo per ciò che attiene al raggiungimento degli obiettivi legati alla produzione di energia da fonti rinnovabili, soprattutto grazie al contributo delle politiche di incentivazione nazionale attuate nel triennio 2011-2013.

Il trend di produzione dalle centrali termoelettriche, per effetto dell'entrata in esercizio di impianti alimentati da fonti rinnovabili, mostra una costante riduzione dal 2009. Il ruolo delle centrali termoelettriche continua, comunque, ad essere rilevante e fondamentale per mantenere in equilibrio lo stato attuale della rete, compensando gli squilibri dovuti alla natura discontinua della produzione rinnovabile (eolica e fotovoltaica).

In realtà le rinnovabili in Sicilia, pur avendo registrato una certa espansione, ancora oggi occupano posizioni secondarie (sesto posto per numero di impianti, quinto posto per potenza installata e nono posto per produzione) se inquadrare nel contesto nazionale. Il trend di crescita, sia del numero d'impianti che della potenza installata, ha subito una inversione di tendenza a partire dal 2013.

I consumi di energia elettrica risultano in contrazione in tutti i settori, verosimilmente legata alla crisi economica.

Nel 2014 l'energia elettrica prodotta da FER in Sicilia si aggira intorno al 23% del fabbisogno finale. Sono, di per sé, eloquenti i quadri consuntivi elaborati da Terna S.p.A., tanto quello relativo alle potenze installate quanto quello riportante le produzioni ed i consumi (Figure 11.1 ed 11.2).

Il risultato ampiamente positivo è, tuttavia, conseguenza delle politiche nazionali di incentivazione (conto energia FV, certificati verdi) e semplificazione amministrativa (D.lgs. 387/03) del decennio scorso. Conclusa la tumultuosa fase di installazione di impianti eolici e fotovoltaici conseguente ai “generosi” incentivi, lo sviluppo degli impianti a FER non risulta “sostenibile” dal punto di vista territoriale e sociale. Si pensi, a talune problematiche di impatto ambientale, alla extraterritorialità degli investimenti e dei ritorni economici, al costo dell'incentivazione presente sulle bollette elettriche, alle problematiche occupazionali degli addetti ai lavori di progettazione, agli iter autorizzativi o di realizzazione degli impianti, ad un costo dell'energia più elevato in Sicilia rispetto a quello di altre regioni, a causa di una rete elettrica vetusta ed incompleta.

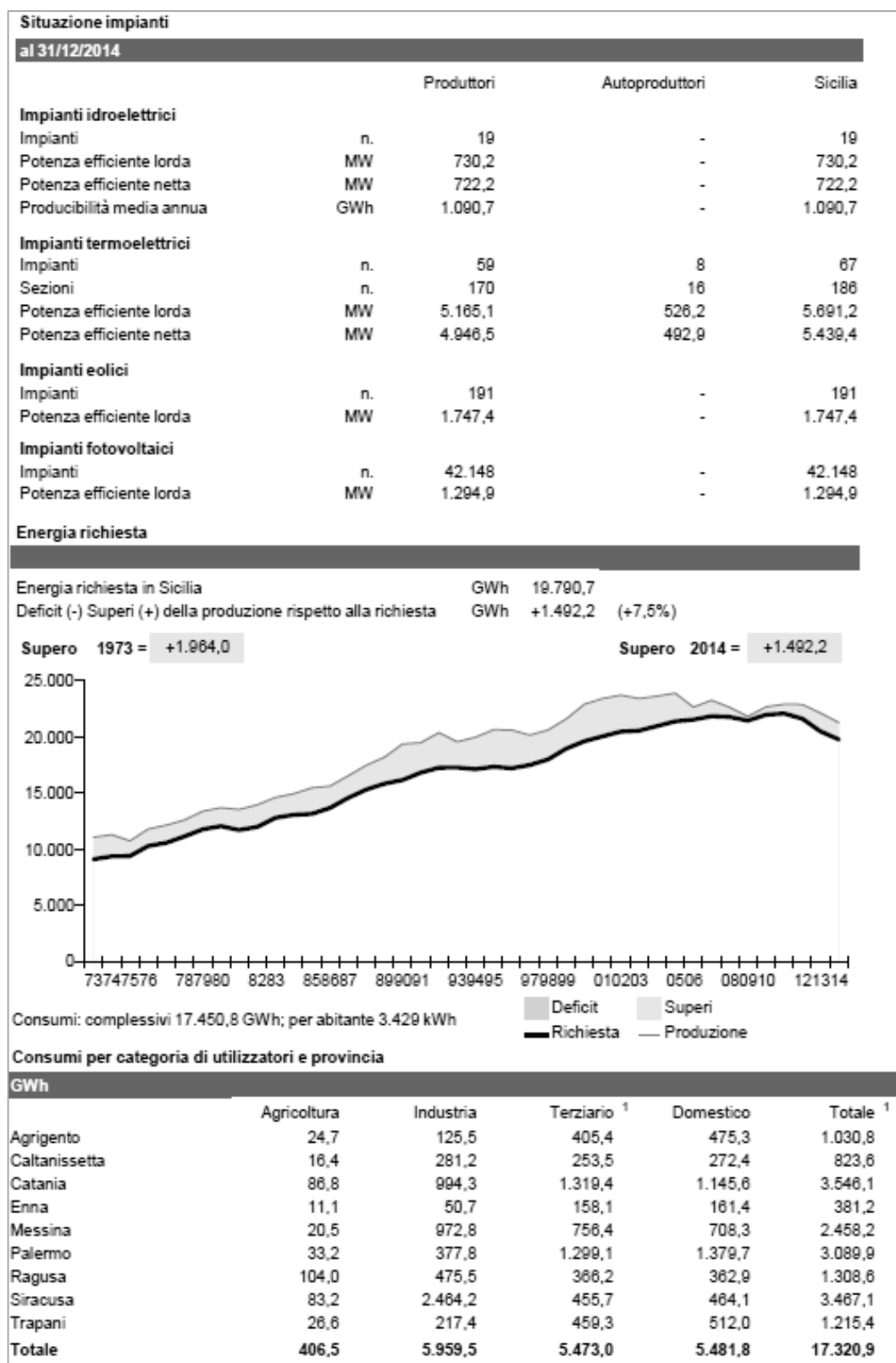


Figura 11.1 - Bilancio energia elettrica Sicilia – potenze installate e consumi.

| Bilancio dell'energia elettrica                     |  |                 |                 |
|---|--|-----------------|-----------------|
| GWh   |  |                 | 2014            |
|   | Operatori del mercato elettrico <sup>2</sup> | Autoproduttori  | Sicilia         |
| <b>Produzione lorda</b>                             |  |                 |                 |
| - idroelettrica                                     | 471,1  | -               | 471,1           |
| - termoelettrica tradizionale                       | 15.377,9                                     | 1.871,3         | 17.249,2        |
| - geotermoelettrica                                 | -  | -               | -               |
| - eolica  | 2.922,4                                      | -               | 2.922,4         |
| - fotovoltaica                                      | 1.893,3                                      | -               | 1.893,3         |
| <b>Totale produzione lorda</b>                      | <b>20.664,7</b>                              | <b>1.871,3</b>  | <b>22.536,1</b> |
|   | -  | -               | -               |
| <b>Servizi ausiliari della Produzione</b>           | <b>622,0</b>                                 | <b>204,2</b>    | <b>826,2</b>    |
|   | =  | =               | =               |
| <b>Produzione netta</b>                             |  |                 |                 |
| - idroelettrica                                     | 460,7  | -               | 460,7           |
| - termoelettrica tradizionale                       | 14.833,1                                     | 1.667,2         | 16.500,3        |
| - geotermoelettrica                                 | -  | -               | -               |
| - eolica  | 2.898,8                                      | -               | 2.898,8         |
| - fotovoltaica                                      | 1.850,1                                      | -               | 1.850,1         |
| <b>Totale produzione netta</b>                      | <b>20.042,7</b>                              | <b>1.667,2</b>  | <b>21.709,8</b> |
|   | -  | -               | -               |
| <b>Energia destinata ai pompaggi</b>                | <b>426,9</b>                                 | -               | <b>426,9</b>    |
|   | =  | =               | =               |
| <b>Produzione destinata al consumo</b>              | <b>19.615,7</b>                              | <b>1.667,2</b>  | <b>21.282,9</b> |
|   | +  | +               | +               |
| <b>Cessioni degli Autoproduttori agli Operatori</b> | <b>+318,2</b>                                | <b>-318,2</b>   | -               |
|   | +  | +               | +               |
| <b>Saldo import/export con l'estero</b>             | -  | -               | -               |
|   | +  | +               | +               |
| <b>Saldo con le altre regioni</b>                   | <b>-1.492,2</b>                              | -               | <b>-1.492,2</b> |
|   | =  | =               | =               |
| <b>Energia richiesta</b>                            | <b>18.441,7</b>                              | <b>1.349,0</b>  | <b>19.790,7</b> |
|   | -  | -               | -               |
| <b>Perdite</b>                                      | <b>2.339,9</b>                               | <b>0,0</b>      | <b>2.339,9</b>  |
|   | =  | =               | =               |
| <b>Consumi</b>                                      | Autoconsumo                                  | 247,7           | 1.340,3         |
|   | Mercato libero <sup>3</sup>                  | 10.538,7        | 8,7             |
|   | Mercato tutelato                             | 5.315,5         | -               |
|   | <b>Totale Consumi</b>                        | <b>16.101,9</b> | <b>1.349,0</b>  |
|   |  |                 | 17.450,8        |

Figura 11.2 - Bilancio energia elettrica Sicilia – produzione e perdite.

In Sicilia la potenza lorda totale degli impianti è di 8.937,60 MW, circa 743,6 MW in meno rispetto al 2013. La figura 11.1 mostra il confronto con la potenza lorda installata in Italia.

|                          | SICILIA            | ITALIA             |
|--------------------------|--------------------|--------------------|
|                          | Potenza lorda [MW] | Potenza lorda [MW] |
| Idroelettrici            | 730,20             | 22.311,70          |
| Termoelettrici           | 5.165,10           | 71.653,50          |
| di cui geotermoelettrici | -                  | 821,00             |
| Eolici                   | 1.747,40           | 8.703,10           |
| Fotovoltaici             | 1.294,90           | 18.609,40          |
| <b>TOTALE</b>            | <b>8.937,60</b>    | <b>122.098,70</b>  |

Tabella 11.1 - Potenza installata Sicilia – Italia anno 2014.



Per quanto riguarda invece la produzione totale lorda nazionale del 2014, questa è stata quantificata da TERNA in 279.828,5 GWh, in diminuzione rispetto al 2013 (289.803,2 GWh), mentre la produzione lorda da fonti rinnovabili (idrica, eolica, fotovoltaica, geotermica e bioenergie) è aumentata del 7,2% rispetto al 2013, raggiungendo i 120.678,9 GWh.

La produzione termoelettrica del 2014, in Italia, rappresenta il 65%, della produzione lorda totale di energia elettrica.

In Sicilia, la produzione lorda è stata di 22.536,1 GWh (netta 21.709,8 GWh) a fronte di una richiesta di 19.790,7 GWh, con un saldo in uscita di 1.492,2 GWh.

La produzione regionale è attribuibile per il 76,6% ad impianti termoelettrici, mentre la produzione da fonti rinnovabili ha chiuso il 2014 con 5.221,3 GWh (23% della produzione regionale).

La Figura 11.3 mostra la produzione lorda di energia elettrica per l'anno 2014 in Sicilia per tipologia.

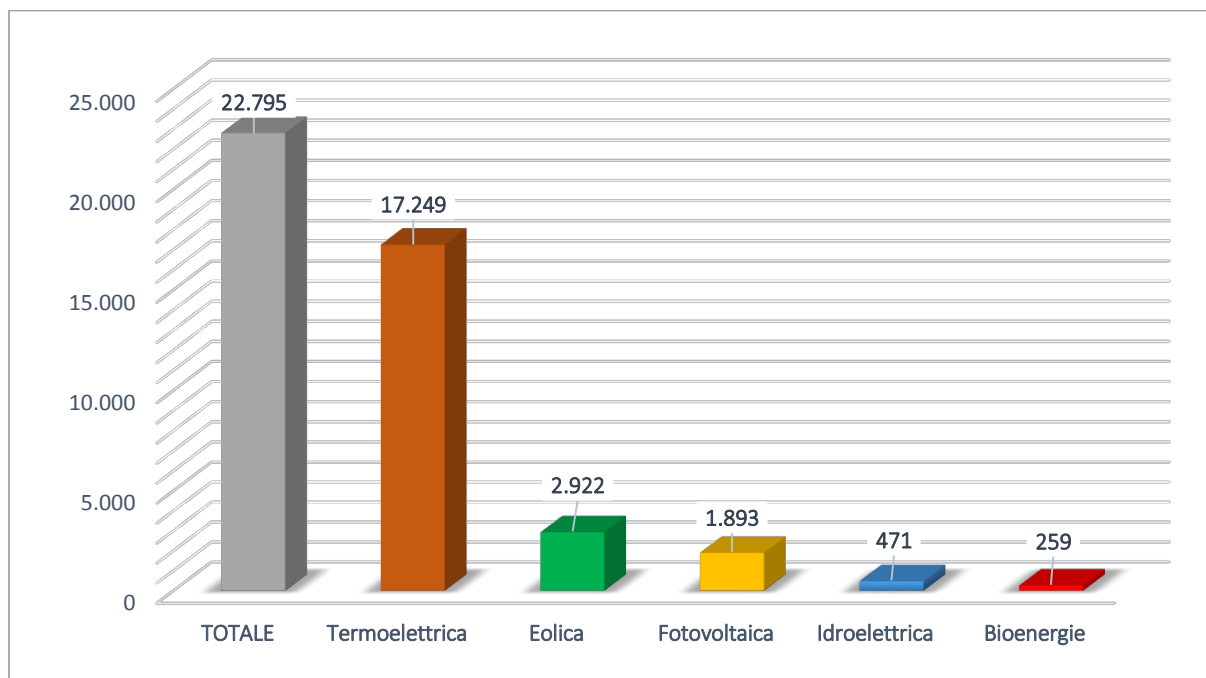


Figura 11.3 - Produzione lorda per tipologia di impianti – Sicilia 2014 [GWh].

Come è evidente, in Sicilia il settore termoelettrico costituisce la maggiore produzione, seguito dall'eolico, mentre a livello nazionale il termoelettrico è seguito dall'idroelettrico.

Dalla figura 11.4 è possibile rilevare il trend decrescente di produzione elettrica da termoelettrico, che come anzidetto è dovuto essenzialmente ad una maggiore produzione da impianti FER-E e da una riduzione dei consumi.

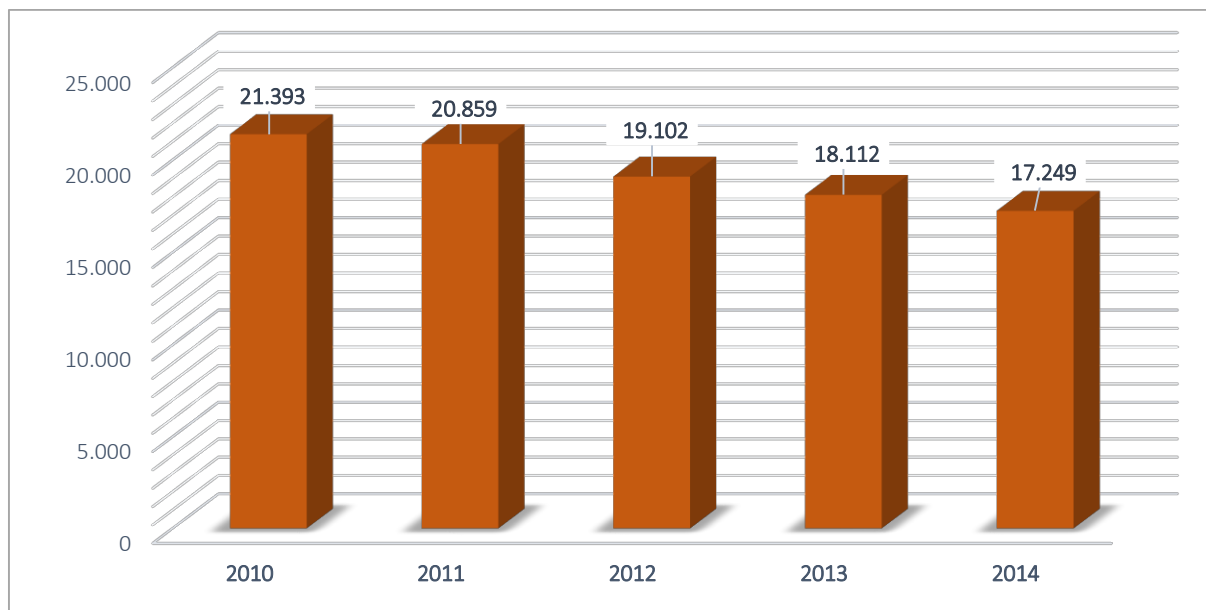


Figura 11.4 - Produzione lorda da termoelettrico – Sicilia 2014 [GWh].

In contrapposizione al trend del termoelettrico, le FER eolico e fotovoltaico nel corso degli anni hanno mostrato un graduale aumento, Figura 11.5, in particolare per il fotovoltaico nel biennio 2011-2012 a causa del mutamento degli strumenti di incentivazione varato a livello nazionale.

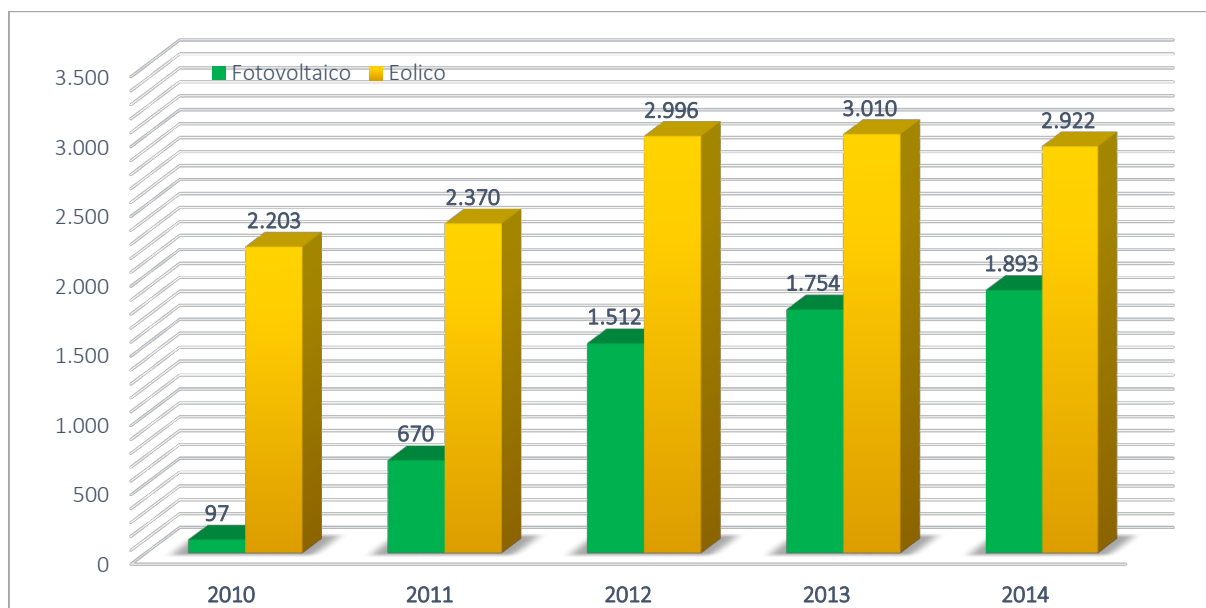


Figura 11.5 - Produzione lorda da eolico e fotovoltaico – Sicilia 2014 [GWh].

L'energia totale richiesta in Sicilia per l'anno 2014 è stata di 17,45 TWh, in calo del 3,6% rispetto al 2013. Questo valore è al netto delle perdite per produzione e trasporto, dell'energia necessaria ai pompaggi e dell'export (Figura 11.6), che è presente con valori rilevanti poiché la sicurezza del sistema elettrico regionale è mantenuta gestendo prevalentemente l'isola in esportazione. Nel complesso questi tre fattori impegnano circa il 19,6% della produzione lorda.

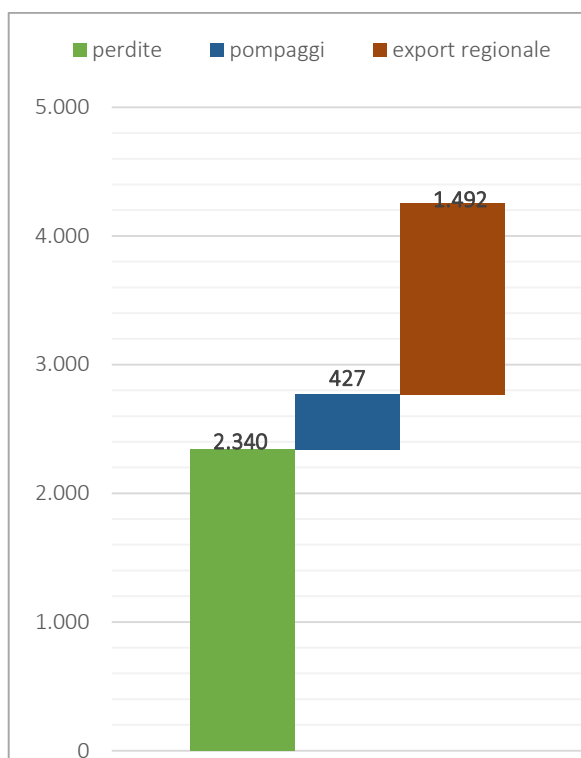


Figura 11.6 - Perdite ed export – Sicilia 2014 [GWh].

La ripartizione dei consumi per macrosettori, Figura 11.7, vede: il settore industriale con un peso del 34,2%, il settore domestico con 31,4% il terziario e l'agricoltura rappresentano rispettivamente il 31,4% ed il 2,3%. Il settore ferroviario rappresenta appena l'1%,

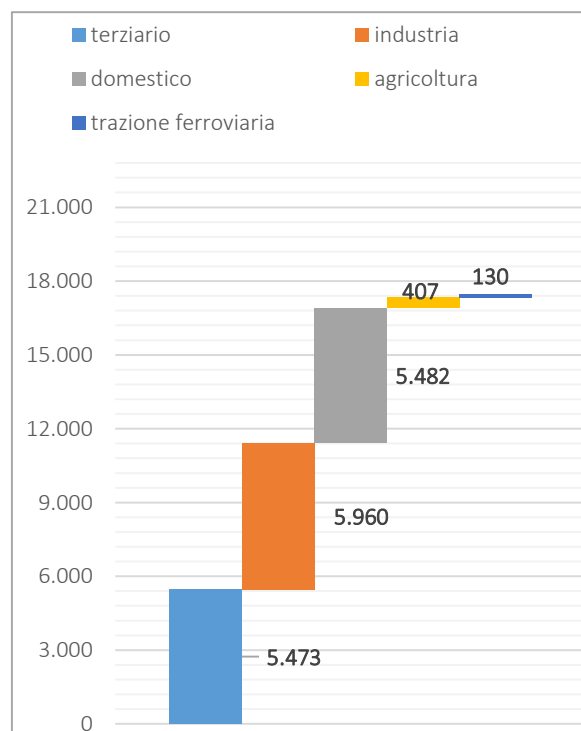


Figura 11.7 - Consumi – Sicilia 2014 [GWh].

L'energia prodotta, Figura 11.8 nell'isola è generata per il 76% agli impianti termoelettrici, per il 13,4% agli impianti eolici, per l'8,5 % ad impianti fotovoltaici ed appena il 2,1% per impianti idroelettrici.

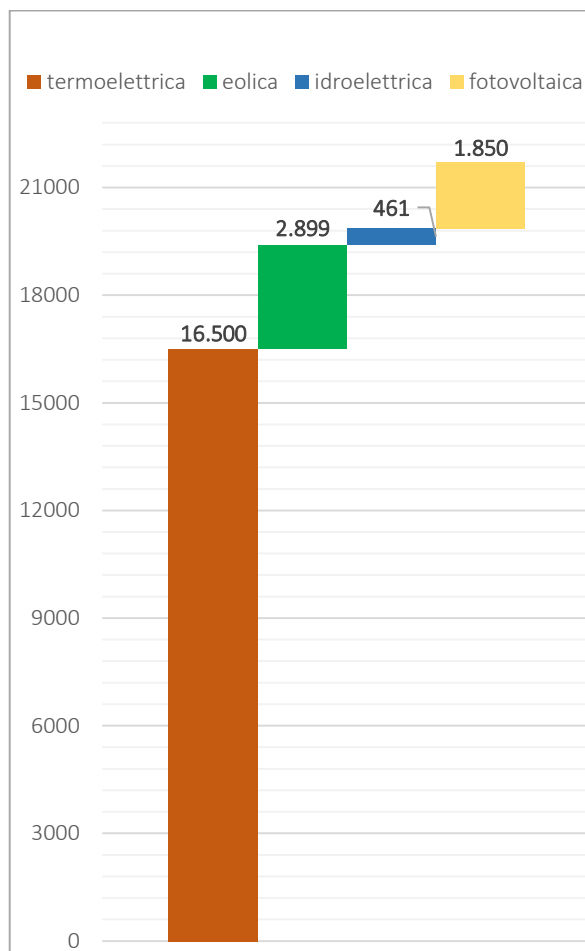


Figura 11.8 - Produzione – Sicilia 2014 [GWh].

Dal confronto con le altre regioni italiane, la Sicilia è al 5° posto come numero di impianti (Figura 11.9), scavalcando nel 2014 la Puglia, 5° per potenza (Figura 11.10) e 8° per produzione (Figura 11.11), superando nel 2014 la Calabria. Se invece si rapporta la produzione da FER-E all'estensione territoriale regionale, la Sicilia occupa il 17° posto (Figura 11.12)

A dicembre 2014, la potenza efficiente lorda installata in Italia da impianti alimentati da FER è stata pari a 50.594,6 MW, più 1,4% rispetto al 2013.

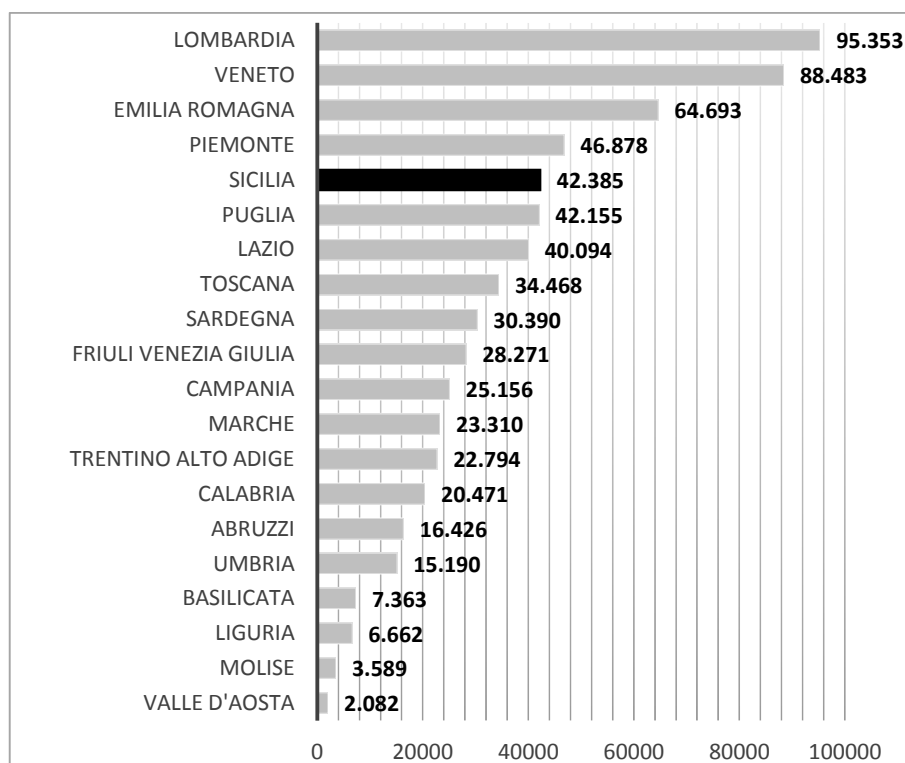


Figura 11.9 - Numero impianti FER – Italia 2014.

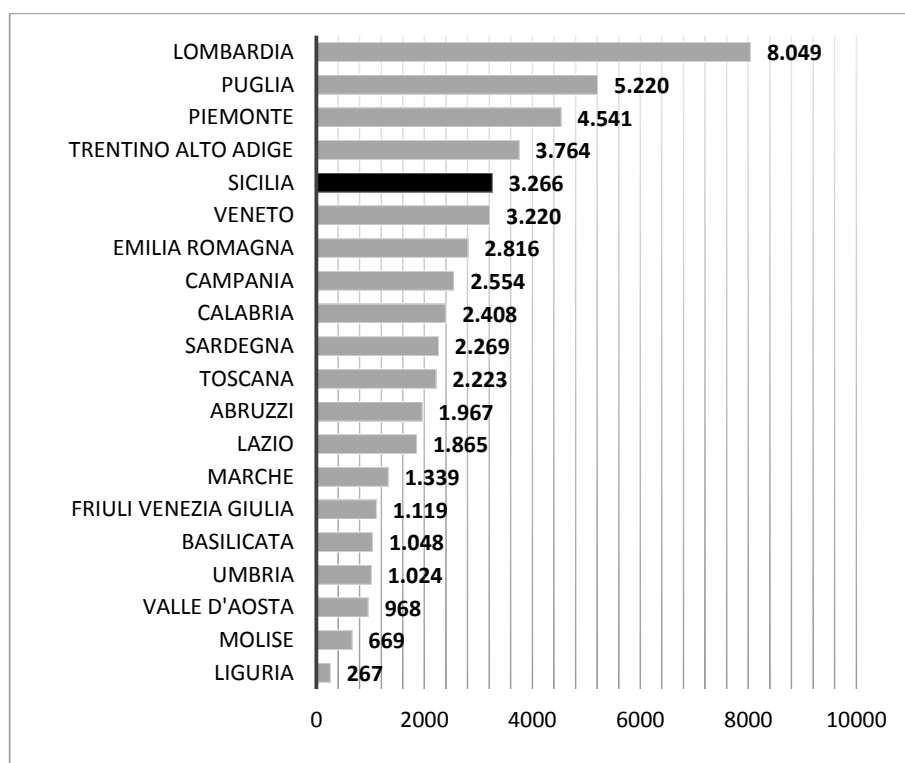


Figura 11.10 - Potenza impianti FER – Italia 2014 [MW].

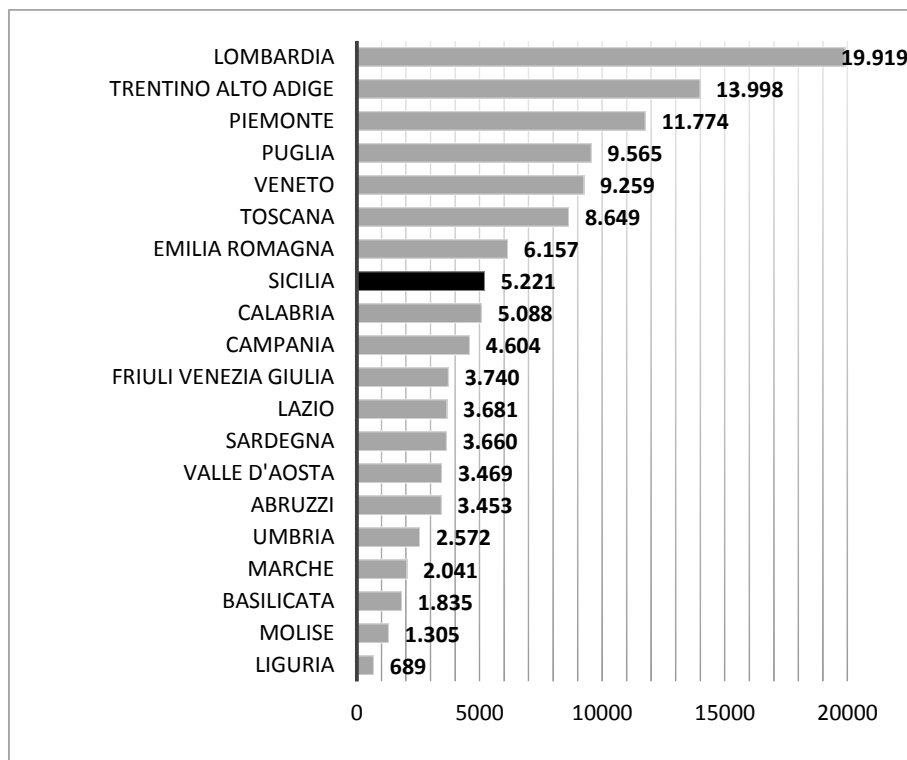


Figura 11.11 - Produzione impianti FER – Italia 2014 [GWh].

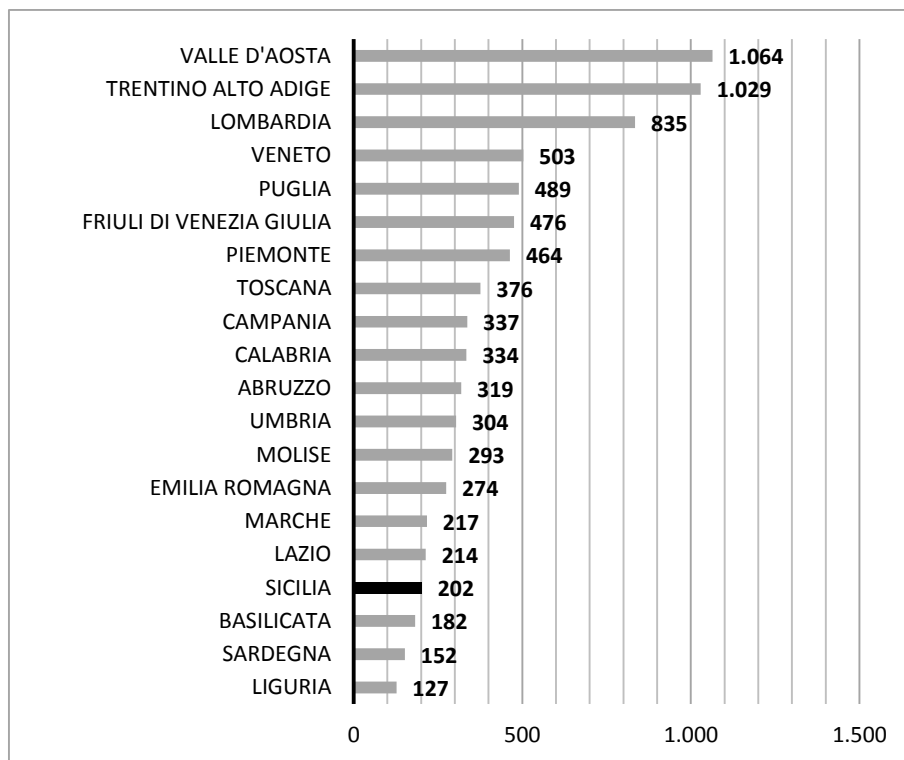


Figura 11.12 - Produzione impianti FER per superficie territoriale – Italia 2014 [MWh/km²].

Dalle analisi per la regione Sicilia, mostrate nelle figure 11.13, 11.14 ed 11.15, è evidente una crescita graduale delle FER-E fino al 2010, poi nel biennio 2011-2012 il boom delle rinnovabili, a cui come

detto ha contribuito il sistema incentivante nazionale soprattutto per il fotovoltaico, e poi una nuova fase di crescita lenta quasi stagnante che presumibilmente perdurerà nei prossimi anni, a meno di una nuova spinta promotrice diretta dalle politiche energetico-economiche settoriali.

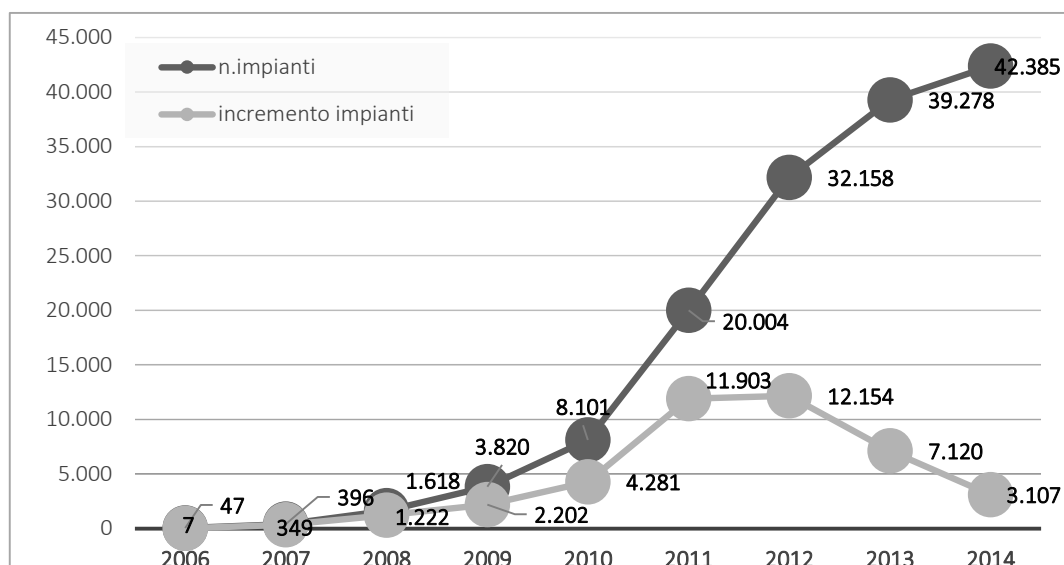


Figura 11.13 - Numero ed incremento impianti FER – Sicilia.

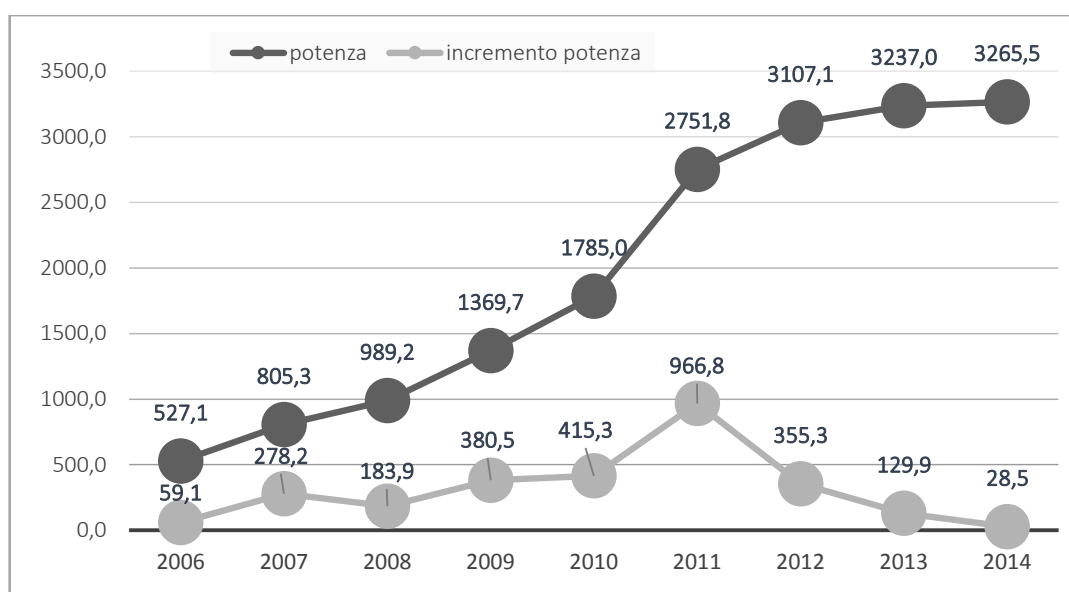


Figura 11.14 - Potenza annuale ed incremento impianti FER – Sicilia [MWh].

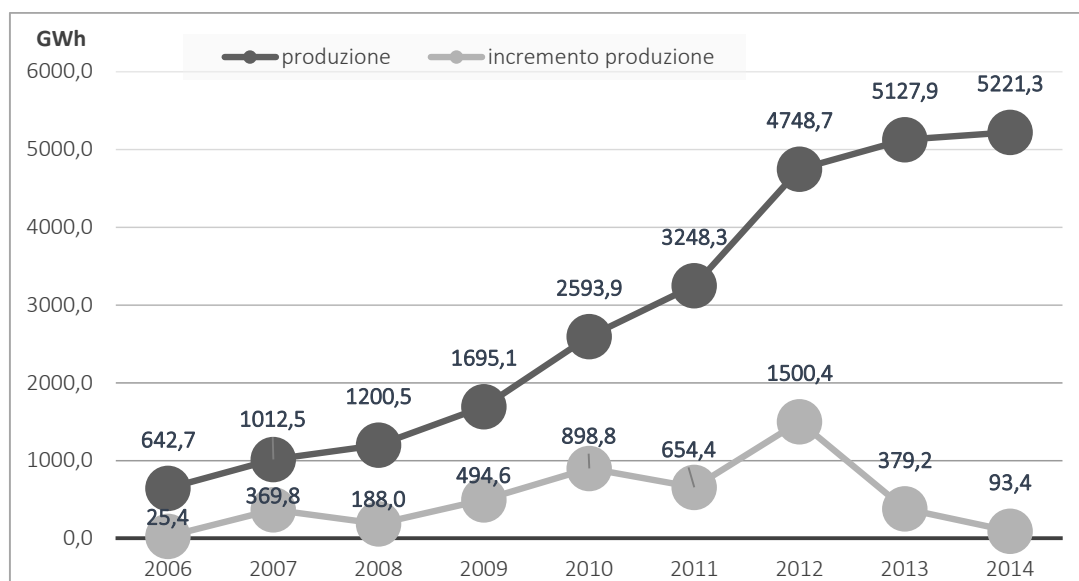


Figura 11.15 - Produzione annuale ed incremento impianti FER – Sicilia.

Per quanto riguarda le singole fonti, la Tabella 11.2 e la rispettiva figura 11.16 mostrano il trend della potenza installata dal 2006 al 2014.

|                      | 2006         | 2007         | 2008         | 2009          | 2010          | 2011          | 2012          | 2013          | 2014          |
|----------------------|--------------|--------------|--------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Bioenergie           | 16,1         | 17,8         | 19,0         | 25,4          | 42,2          | 53,9          | 80,8          | 80,4          | 73,0          |
| Eolico               | 358,6        | 630,9        | 794,6        | 1147,9        | 1435,6        | 1680,9        | 1749,2        | 1750,2        | 1747,4        |
| Fotovoltaico         | 0,2          | 4,4          | 17,4         | 44,2          | 155,9         | 865,7         | 1125,8        | 1256,2        | 1294,9        |
| Idrico               | 152,2        | 152,2        | 152,2        | 152,2         | 151,3         | 151,3         | 151,3         | 150,1         | 150,2         |
| <b>POTENZA FER-E</b> | <b>527,1</b> | <b>805,3</b> | <b>983,2</b> | <b>1369,7</b> | <b>1785,0</b> | <b>2751,8</b> | <b>3107,1</b> | <b>3236,9</b> | <b>3265,5</b> |

Tabella 11.2 - Potenza impianti FER – Sicilia [MW].

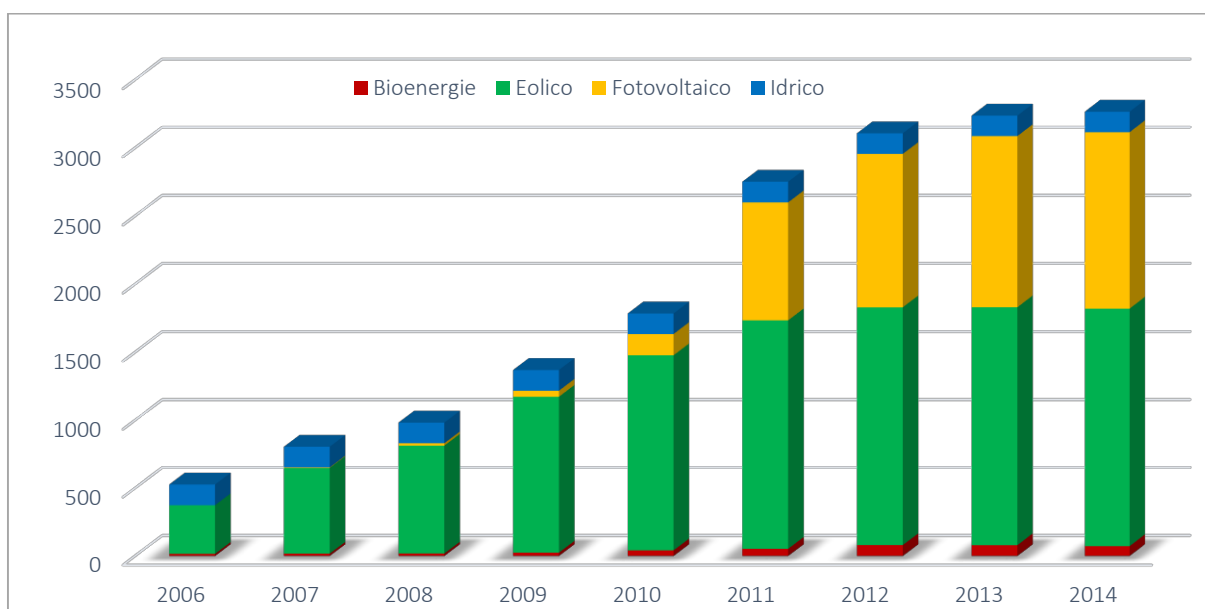


Figura 11.16 - Potenza impianti FER – Sicilia [MW].



Nella Tabella 11.3 e nella rispettiva figura 11.17 è mostrato il trend dell'energia prodotta per singola fonte dal 2006 al 2014.

|                     | 2006  | 2007  | 2008   | 2009   | 2010   | 2011   | 2012   | 2013   | 2014   |
|---------------------|-------|-------|--------|--------|--------|--------|--------|--------|--------|
| <b>Bioenergie</b>   | 62,9  | 58,7  | 75,5   | 113,6  | 150,2  | 109,8  | 69,6   | 189,8  | 259,2  |
| <b>Eolico</b>       | 488,7 | 854,7 | 1044,0 | 1444,4 | 2203,0 | 2369,9 | 2995,9 | 3009,5 | 2922,4 |
| <b>Fotovoltaico</b> | 0,1   | 1,5   | 10,7   | 33,3   | 97,2   | 670,4  | 1511,5 | 1754,0 | 1893,3 |
| <b>Idrico</b>       | 91,0  | 97,5  | 70,3   | 103,8  | 143,6  | 98,1   | 171,7  | 174,7  | 146,4  |

Tabella 11.3 - Produzione impianti FER – Sicilia [GWh].

Come è da attendersi, rivestono importanza per GWh prodotti le fonti eolica e fotovoltaica. Le fonti di bio-energia e l'idroelettrico rimangono a tutt'oggi marginali nel panorama energetico siciliano. L'idroelettrico, in particolare a causa del potenziale tecnico siciliano, legato fortemente all'orografia del territorio e alle caratteristiche geoclimatiche, è molto basso rispetto ad altre regioni del nord Italia.

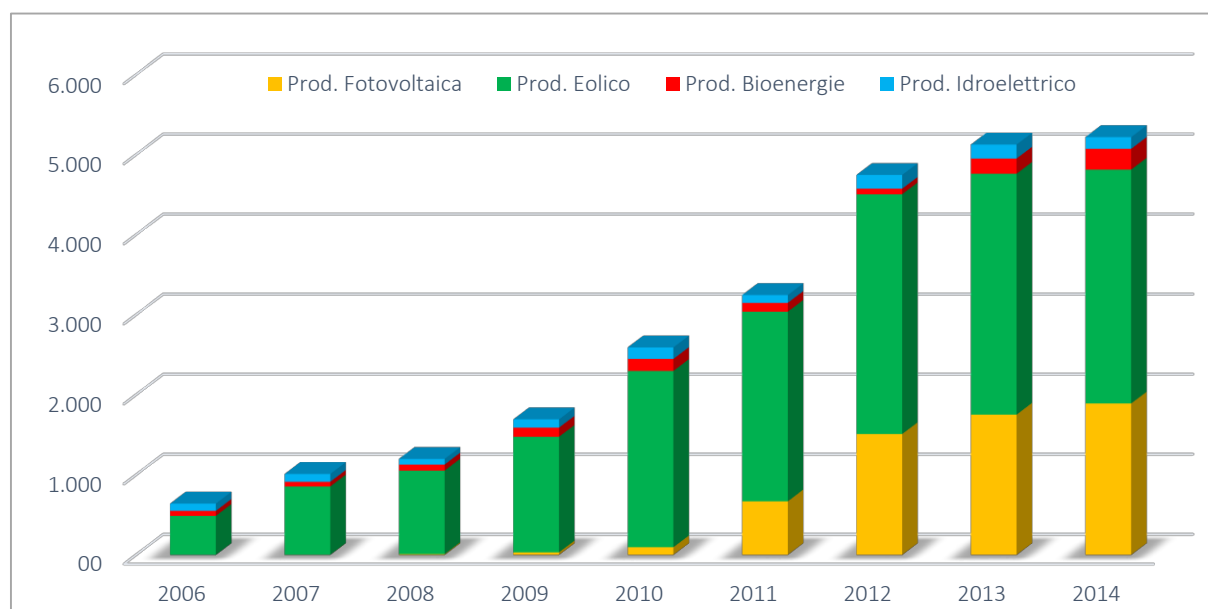


Figura 11.17 - Produzione impianti FER – Sicilia [GWh].

## 11.2 I dati energetici

La Costruzione del DataSet dei dati di produzione del FER-E per la regione Sicilia è stato svolto grazie ai report annuali pubblicati dai soggetti nazionali responsabili della trasmissione dell'energia elettrica, Terna S.p.A. [38] e della qualificazione dell'energia da FER, GSE [39-42]. Tali soggetti forniscono annualmente i dati relativi all'anno precedente, pertanto è stato possibile costruire un DataSet che vada dal 2011 al 2014. Poiché i dati forniti dal GSE sono giornalieri mentre quelli forniti da TERNA sono orari, questi ultimi sono stati riportati a dati giornalieri.

Si riporta di seguito campione del DataSet realizzato (Tabella 14.4).

| Data       | Idro_fl_tep | PV_tep | Eolico_tep | kW_PV     | kW_Eolico  | GW_FER  |
|------------|-------------|--------|------------|-----------|------------|---------|
| 01/01/2011 | 0           | 8.77   | 182.06     | 157844.66 | 1436272.05 | 1787.65 |
| 02/01/2011 | 0           | 21.67  | 104.32     | 159789.32 | 1436944.11 | 1790.30 |
| 03/01/2011 | 0           | 46.53  | 53.49      | 161733.97 | 1437616.16 | 1792.95 |
| 04/01/2011 | 0           | 1.38   | 64.33      | 163678.63 | 1438288.22 | 1795.60 |
| 05/01/2011 | 0           | 23.31  | 59.60      | 165623.29 | 1438960.27 | 1798.24 |
| 06/01/2011 | 0           | 30.70  | 249.83     | 167567.95 | 1439632.33 | 1800.89 |
| 07/01/2011 | 0           | 31.22  | 627.37     | 169512.60 | 1440304.38 | 1803.54 |
| 08/01/2011 | 0           | 47.39  | 435.85     | 171457.26 | 1440976.44 | 1806.19 |
| 09/01/2011 | 0           | 35.00  | 273.14     | 173401.92 | 1441648.49 | 1808.84 |
| 10/01/2011 | 0           | 33.11  | 516.09     | 175346.58 | 1442320.55 | 1811.49 |
| 11/01/2011 | 0           | 44.29  | 339.36     | 177291.23 | 1442992.60 | 1814.14 |
| 12/01/2011 | 0           | 28.29  | 218.35     | 179235.89 | 1443664.66 | 1816.79 |
| 13/01/2011 | 0           | 47.56  | 517.55     | 181180.55 | 1444336.71 | 1819.43 |
| 14/01/2011 | 0           | 52.98  | 195.48     | 183125.21 | 1445008.77 | 1822.08 |
| 15/01/2011 | 0           | 54.35  | 511.61     | 185069.86 | 1445680.82 | 1824.73 |
| 16/01/2011 | 0           | 54.95  | 407.98     | 187014.52 | 1446352.88 | 1827.38 |
| 17/01/2011 | 0           | 52.46  | 83.68      | 188959.18 | 1447024.93 | 1830.03 |
| 18/01/2011 | 0           | 41.88  | 123.07     | 190903.84 | 1447696.99 | 1832.68 |
| 19/01/2011 | 0           | 28.38  | 36.98      | 192848.49 | 1448369.04 | 1835.33 |
| 20/01/2011 | 0           | 31.91  | 204.08     | 194793.15 | 1449041.10 | 1837.98 |
| 21/01/2011 | 0           | 48.07  | 439.98     | 196737.81 | 1449713.15 | 1840.62 |
| 22/01/2011 | 0           | 54.18  | 1237.88    | 198682.47 | 1450385.21 | 1843.27 |
| 23/01/2011 | 0           | 37.07  | 413.23     | 200627.12 | 1451057.26 | 1845.92 |
| 24/01/2011 | 0           | 36.29  | 345.29     | 202571.78 | 1451729.32 | 1848.57 |
| 25/01/2011 | 0           | 45.41  | 137.08     | 204516.44 | 1452401.37 | 1851.22 |
| 26/01/2011 | 0           | 15.14  | 129.95     | 206461.10 | 1453073.42 | 1853.87 |
| 27/01/2011 | 0           | 1.72   | 626.68     | 208405.75 | 1453745.48 | 1856.52 |

Tabella 11.4 - DataSet energetico.

### 11.3 I dati meteo climatici

Per quanto riguarda i dati meteo della Sicilia è stata svolta una ricognizione dei dati presenti in diversi archivi regionali (SIAS) nazionali (Aeronautica Militare), verificando al contempo la possibilità di allineamento dei dati climatici con i dati energetici presenti nel DataSet.

Si è quindi deciso di prendere i dati relativi alle stazioni meteorologiche site nella provincia di Palermo, Trapani e Catania. Ed in particolare presso l'Aeroporto di Punta Raisi (PA), l'Aeroporto di Trapani, e l'Aeroporto di Catania.

Il DataSet climatico utilizzato è riportato nella seguente Tabella 14.5.

| DATA       | T_MEDIA<br>[°C] | T_MIN<br>[°C] | T_MAX<br>[°C] | PUNTO_RUGIADA<br>[°C] | UMIDITA<br>[%] | VISIBILITA<br>[km] | VENTO_MEDIA<br>[km/h] | VENTO_MAX<br>[km/h] | RAFFICA<br>[km/h] |
|------------|-----------------|---------------|---------------|-----------------------|----------------|--------------------|-----------------------|---------------------|-------------------|
| 01/01/2011 | 12              | 9             | 15            | 7                     | 71             | 11                 | 4                     | 9                   | 0                 |
| 02/01/2011 | 12              | 8             | 14            | 7                     | 74             | 10                 | 4                     | 8                   | 0                 |
| 03/01/2011 | 11              | 8             | 15            | 7                     | 74             | 9                  | 5                     | 10                  | 0                 |
| 04/01/2011 | 11              | 9             | 14            | 8                     | 81             | 10                 | 5                     | 10                  | 0                 |
| 05/01/2011 | 12              | 9             | 14            | 6                     | 67             | 10                 | 12                    | 100                 | 0                 |
| 06/01/2011 | 13              | 10            | 16            | 7                     | 62             | 10                 | 7                     | 13                  | 0                 |
| 07/01/2011 | 14              | 10            | 18            | 8                     | 65             | 16                 | 5                     | 10                  | 0                 |

|            |    |    |    |   |    |    |    |    |    |
|------------|----|----|----|---|----|----|----|----|----|
| 08/01/2011 | 14 | 10 | 19 | 8 | 66 | 14 | 4  | 8  | 0  |
| 09/01/2011 | 14 | 10 | 17 | 7 | 65 | 13 | 4  | 12 | 0  |
| 10/01/2011 | 16 | 12 | 19 | 8 | 58 | 15 | 11 | 20 | 23 |
| 11/01/2011 | 13 | 11 | 14 | 9 | 75 | 10 | 10 | 13 | 0  |
| 12/01/2011 | 12 | 10 | 14 | 8 | 74 | 10 | 10 | 24 | 0  |
| 13/01/2011 | 13 | 10 | 15 | 8 | 68 | 11 | 10 | 21 | 0  |
| 14/01/2011 | 11 | 8  | 15 | 7 | 76 | 13 | 4  | 7  | 0  |
| 15/01/2011 | 12 | 9  | 15 | 9 | 80 | 11 | 6  | 16 | 0  |
| 16/01/2011 | 13 | 11 | 15 | 9 | 77 | 10 | 13 | 18 | 0  |
| 17/01/2011 | 12 | 9  | 15 | 9 | 79 | 11 | 6  | 10 | 0  |
| 18/01/2011 | 12 | 9  | 18 | 8 | 77 | 10 | 7  | 14 | 0  |
| 19/01/2011 | 12 | 9  | 14 | 6 | 67 | 11 | 5  | 8  | 0  |
| 20/01/2011 | 13 | 11 | 15 | 7 | 68 | 11 | 11 | 17 | 0  |
| 21/01/2011 | 12 | 9  | 13 | 5 | 64 | 11 | 13 | 22 | 0  |
| 22/01/2011 | 10 | 7  | 12 | 5 | 67 | 9  | 23 | 31 | 43 |
| 23/01/2011 | 10 | 7  | 12 | 4 | 65 | 10 | 12 | 22 | 28 |
| 24/01/2011 | 10 | 7  | 11 | 4 | 68 | 11 | 10 | 76 | 18 |

Tabella 11.5 - DataSet climatico.

## 11.4 Il DataSet

Allineando i dati energetici con quelli meteo è stato possibile realizzare il DataSet definitivo per la costruzione del modello predittivo. È stato svolto quindi l'allineamento dei dati ed eliminate le righe dove era presente un valore nullo per dato mancante. Per quanto riguarda i dati meteo è stato introdotto un parametro che è pari alla somma dei dati nei tre punti di riferimento (Palermo, Trapani e Catania) che fornisce riscontro delle condizioni climatiche complessive dell'isola, visto che i dati di produzione sono anch'essi complessivi della regione.

La Tabella 14.6 riporta un campione del data set utilizzato.

| Data       | Idro | PV    | Eolico | kW_PV     | kW_Eolico  | GW_FER  | T_MEDIA [°C] | T_MIN [°C] | T_MAX [°C] | UMIDITA [%] | VISIBILITA [km] | VENTO_MEDIO [km/h] | VENTO_MAX [km/h] | RAFFICA [km/h] |
|------------|------|-------|--------|-----------|------------|---------|--------------|------------|------------|-------------|-----------------|--------------------|------------------|----------------|
| 01/01/2011 | 0    | 8.77  | 182.06 | 157844.66 | 1436272.05 | 1787.65 | 34           | 24         | 43         | 256         | 27              | 14                 | 34               | 0              |
| 02/01/2011 | 0    | 21.67 | 104.32 | 159789.32 | 1436944.11 | 1790.30 | 33           | 22         | 45         | 253         | 30              | 10                 | 24               | 0              |
| 03/01/2011 | 0    | 46.53 | 53.49  | 161733.97 | 1437616.16 | 1792.95 | 29           | 18         | 43         | 246         | 30              | 13                 | 27               | 0              |
| 04/01/2011 | 0    | 1.38  | 64.33  | 163678.63 | 1438288.22 | 1795.60 | 30           | 22         | 39         | 262         | 29              | 11                 | 22               | 0              |
| 05/01/2011 | 0    | 23.31 | 59.60  | 165623.29 | 1438960.27 | 1798.24 | 32           | 24         | 43         | 240         | 29              | 18                 | 113              | 0              |
| 06/01/2011 | 0    | 30.70 | 249.83 | 167567.95 | 1439632.33 | 1800.89 | 35           | 24         | 49         | 237         | 30              | 18                 | 30               | 0              |
| 07/01/2011 | 0    | 31.22 | 627.37 | 169512.60 | 1440304.38 | 1803.54 | 38           | 27         | 53         | 238         | 38              | 22                 | 38               | 0              |
| 08/01/2011 | 0    | 47.39 | 435.85 | 171457.26 | 1440976.44 | 1806.19 | 41           | 29         | 57         | 236         | 39              | 20                 | 35               | 0              |
| 09/01/2011 | 0    | 35.00 | 273.14 | 173401.92 | 1441648.49 | 1808.84 | 38           | 22         | 53         | 227         | 32              | 14                 | 35               | 0              |
| 10/01/2011 | 0    | 33.11 | 516.09 | 175346.58 | 1442320.55 | 1811.49 | 41           | 27         | 56         | 220         | 35              | 24                 | 47               | 23             |
| 11/01/2011 | 0    | 44.29 | 339.36 | 177291.23 | 1442992.60 | 1814.14 | 36           | 24         | 48         | 233         | 33              | 22                 | 40               | 0              |
| 12/01/2011 | 0    | 28.29 | 218.35 | 179235.89 | 1443664.66 | 1816.79 | 33           | 25         | 43         | 248         | 31              | 20                 | 48               | 23             |
| 13/01/2011 | 0    | 47.56 | 517.55 | 181180.55 | 1444336.71 | 1819.43 | 33           | 20         | 46         | 234         | 32              | 20                 | 38               | 0              |
| 14/01/2011 | 0    | 52.98 | 195.48 | 183125.21 | 1445008.77 | 1822.08 | 30           | 16         | 49         | 241         | 38              | 12                 | 23               | 0              |
| 15/01/2011 | 0    | 54.35 | 511.61 | 185069.86 | 1445680.82 | 1824.73 | 32           | 17         | 48         | 251         | 33              | 14                 | 38               | 0              |
| 16/01/2011 | 0    | 54.95 | 407.98 | 187014.52 | 1446352.88 | 1827.38 | 34           | 23         | 49         | 244         | 32              | 24                 | 37               | 0              |
| 17/01/2011 | 0    | 52.46 | 83.68  | 188959.18 | 1447024.93 | 1830.03 | 33           | 19         | 49         | 248         | 35              | 14                 | 25               | 0              |
| 18/01/2011 | 0    | 41.88 | 123.07 | 190903.84 | 1447696.99 | 1832.68 | 32           | 20         | 49         | 246         | 29              | 17                 | 34               | 0              |
| 19/01/2011 | 0    | 28.38 | 36.98  | 192848.49 | 1448369.04 | 1835.33 | 32           | 19         | 44         | 229         | 28              | 12                 | 23               | 0              |
| 20/01/2011 | 0    | 31.91 | 204.08 | 194793.15 | 1449041.10 | 1837.98 | 37           | 30         | 45         | 230         | 30              | 21                 | 44               | 0              |
| 21/01/2011 | 0    | 48.07 | 439.98 | 196737.81 | 1449713.15 | 1840.62 | 33           | 23         | 43         | 215         | 31              | 28                 | 52               | 24             |

Tabella 11.6 - DataSet utilizzato nei modelli di previsione.

## 11.5 Analisi del DataSet

Tramite il software GMDH Shell si è proceduto all'analisi dei dati del DataSet ed i risultati sono riportati nella Tabella

|                      | <b>Idro_fl_tep</b> | <b>PV_tep</b> | <b>Eolico_tep</b> | <b>kW_PV</b> | <b>kW_Eolico</b> | <b>GW_FER</b> |
|----------------------|--------------------|---------------|-------------------|--------------|------------------|---------------|
| Numero vaolri        | 1461               | 1461          | 1461              | 1461         | 1461             | 1461          |
| Valori di testo      | 0                  | 0             | 0                 | 0            | 0                | 0             |
| Valori mancanti      | 0                  | 0             | 0                 | 0            | 0                | 0             |
| Vaoliri unici        | 71                 | 1400          | 1437              | 1461         | 1461             | 1461          |
| Valori nulli         | 1355               | 0             | 0                 | 0            | 0                | 0             |
| Valore più frequente | 0                  | 241,74        | 175,95            |              |                  |               |
| Valore minimo        | 0                  | 1,37          | 35,17             | 157844,65    | 1436272,05       | 1787,64       |
| Valore massimo       | 194                | 685,59        | 2857,26           | 1294900      | 1750200          | 3265,50       |
| Media                | 0                  | 288,27        | 488,05            | 1125800      | 1747469,04       | 3107,10       |
| Valore medio         | 7,30               | 310,98        | 667,58            | 993666,49    | 1693071,83       | 2905,78       |
| Deviazione standard  | 27,68              | 168,040       | 551,86            | 316082,54    | 86993,00         | 413,79        |

Tabella 11.7 - Statistiche e caratteristiche DataSet parte energetica.

|                      | <b>S_T_MEDIA<br/>[°C]</b> | <b>S_T_MIN<br/>[°C]</b> | <b>S_T_MAX<br/>[°C]</b> | <b>S_UMIDITA<br/>[%]</b> | <b>S_VISIBILITA<br/>[km]</b> | <b>S_VENTO_MEDIO<br/>[km/h]</b> | <b>S_VENTO_MAX<br/>[km/h]</b> | <b>S_RAFFICA<br/>[km/h]</b> |
|----------------------|---------------------------|-------------------------|-------------------------|--------------------------|------------------------------|---------------------------------|-------------------------------|-----------------------------|
| Numero vaolri        | 1461                      | 1461                    | 1461                    | 1461                     | 1461                         | 1461                            | 1461                          | 1461                        |
| Valori di testo      | 0                         | 0                       | 0                       | 0                        | 0                            | 0                               | 0                             | 0                           |
| Valori mancanti      | 0                         | 0                       | 0                       | 0                        | 0                            | 0                               | 0                             | 0                           |
| Vaoliri unici        | 75                        | 74                      | 86                      | 128                      | 44                           | 102                             | 144                           | 140                         |
| Valori nulli         | 12                        | 12                      | 12                      | 12                       | 12                           | 12                              | 12                            | 1005                        |
| Valore più frequente | 48                        | 24                      | 57                      | 209                      | 32                           | 36                              | 76                            | 0                           |
| Valore minimo        | 0                         | 0                       | 0                       | 0                        | 0                            | 0                               | 0                             | 0                           |
| Valore massimo       | 93                        | 76                      | 116                     | 274                      | 63                           | 125                             | 187                           | 248                         |
| Media                | 54                        | 41                      | 66                      | 224                      | 32                           | 37                              | 74                            | 0                           |
| Valore medio         | 54,50                     | 41,53                   | 67,15                   | 219,48                   | 32,95                        | 41,03                           | 77,58                         | 26,95                       |
| Deviazione standard  | 18,01                     | 16,91                   | 19,74                   | 31,67                    | 6,49                         | 18,20                           | 28,64                         | 49,00                       |

Tabella 11.7 - Sstatistiche e caratteristiche DataSet parte climatica.

Si è quindi svolta un'analisi delle correlazioni tra i dati del DataSet, in primo luogo è stata creata la matrice degli scatter (Figure 11.18 ed 11.19).

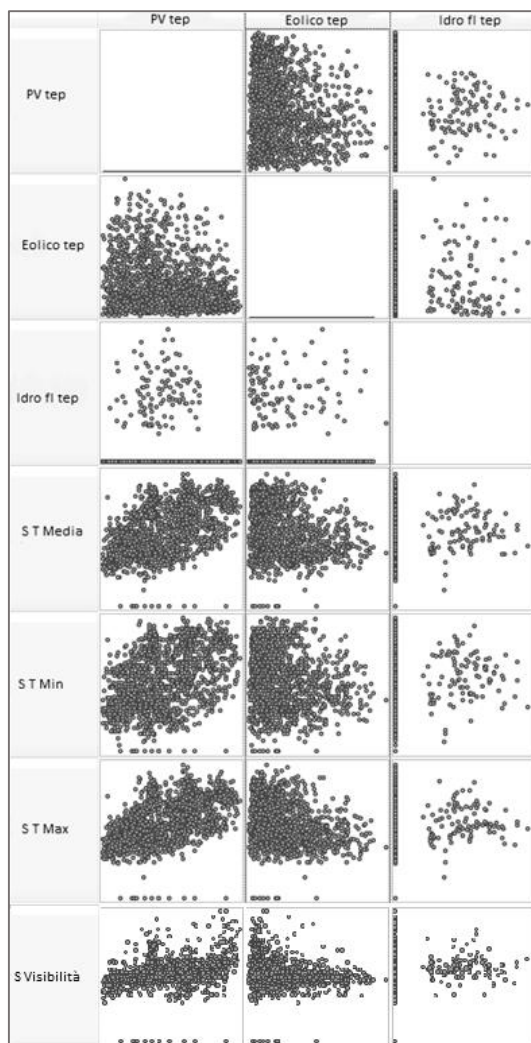


Figura 11.18 - Matrice degli scatter del DataSet.

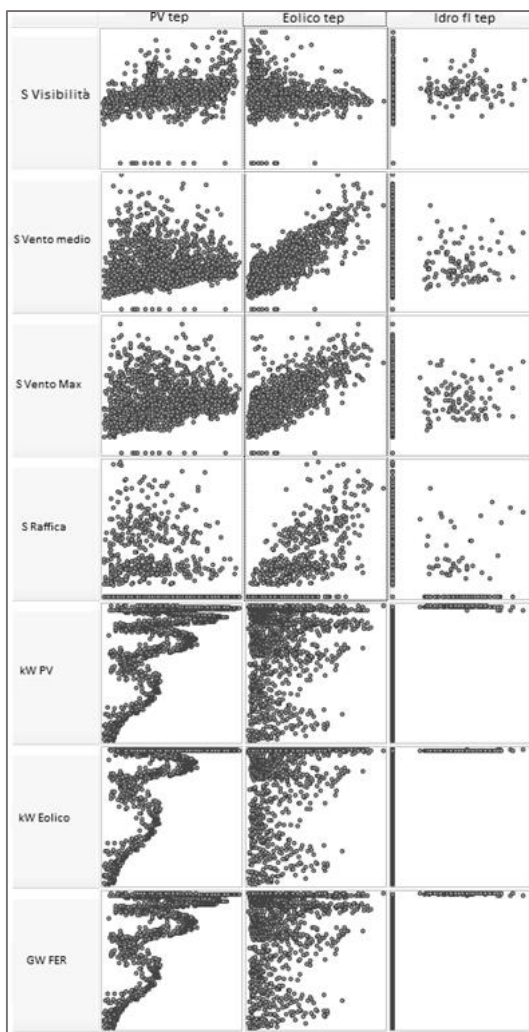


Figura 11.19 - Matrice degli scatter del DataSet.

Per i target è stata quindi verificata la correlazione tramite l'apposito indice (Figure 11.20, 11.21 ed 11.22). Il valore di indice positivo (blu) indica che il valore targhet cresce con il crescere del valore del fattore correlato, indici negativi (rosso) indicano invece che il valore target diminuisce con il crescere del valore del fattore correlato.

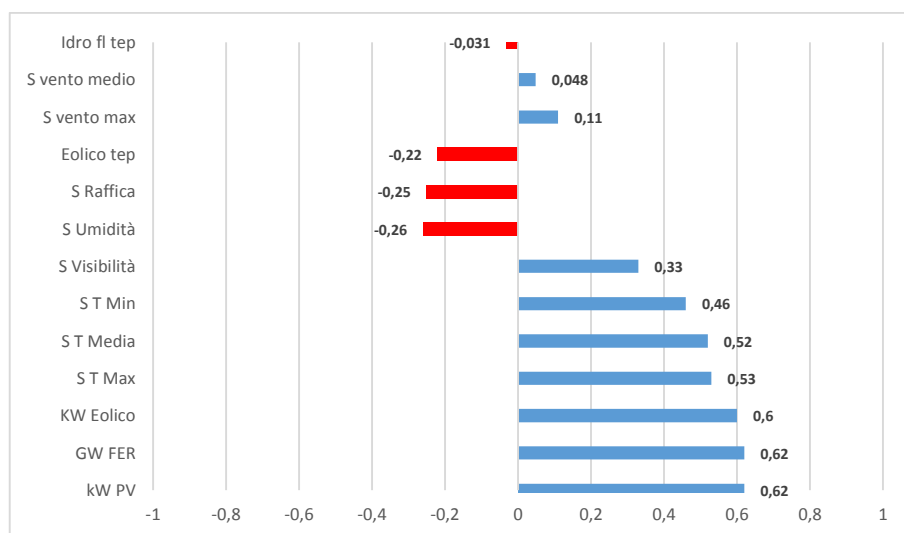


Figura 11.20 - Indici di correlazione per PV tep.

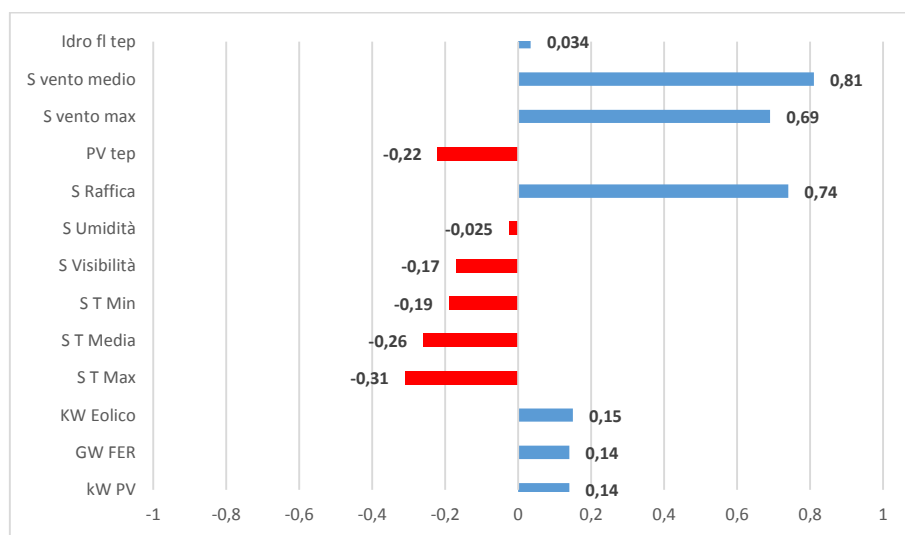


Figura 11.21 - Indici di correlazione per Eolico tep.

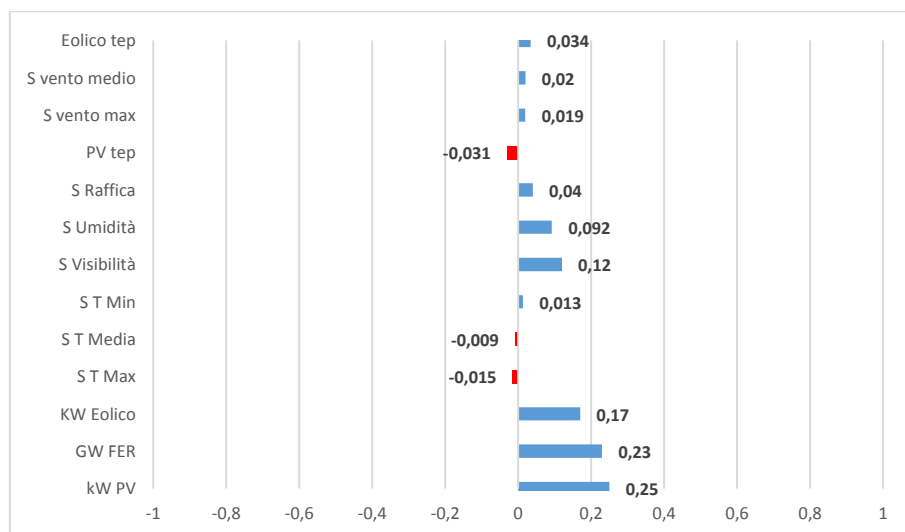


Figura 11.22 - Indici di correlazione per Idro fl tep.

Sempre per i target è stata elaborata graficamente la correlazione con le condizioni climatiche medie di temperatura e ventosità (Figure 11.23, 11.24 ed 11.25).

Dalle analisi realizzate si verifica che i valori di produzione maggiore da fotovoltaico si hanno con indice S\_T\_Media elevato ed S\_Vento\_Medio basso, mentre per la produzione da eolico l'indice S\_T\_Media deve essere compreso tra 20 e 60 °C mentre l'indice S-Vento\_Medio deve essere elevato.

Per quanto riguarda la produzione da idroelettrico, questa non sembra essere collegata a valori specifici dei due indici, come anche evidente dalla Figura 11.25.

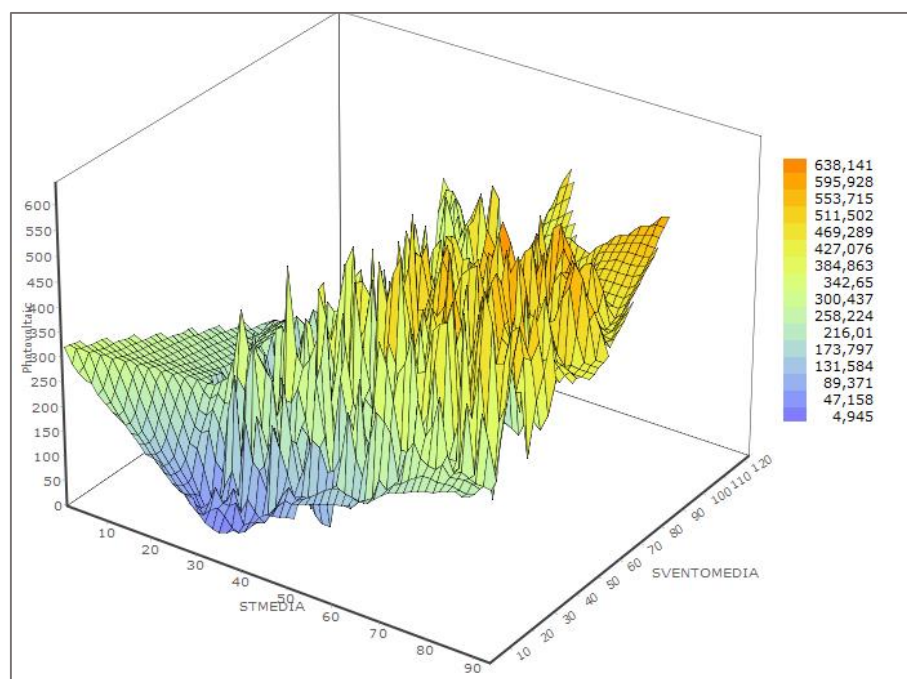


Figura 11.23 - Correlazione tra PV tep ed indici climatici medi.

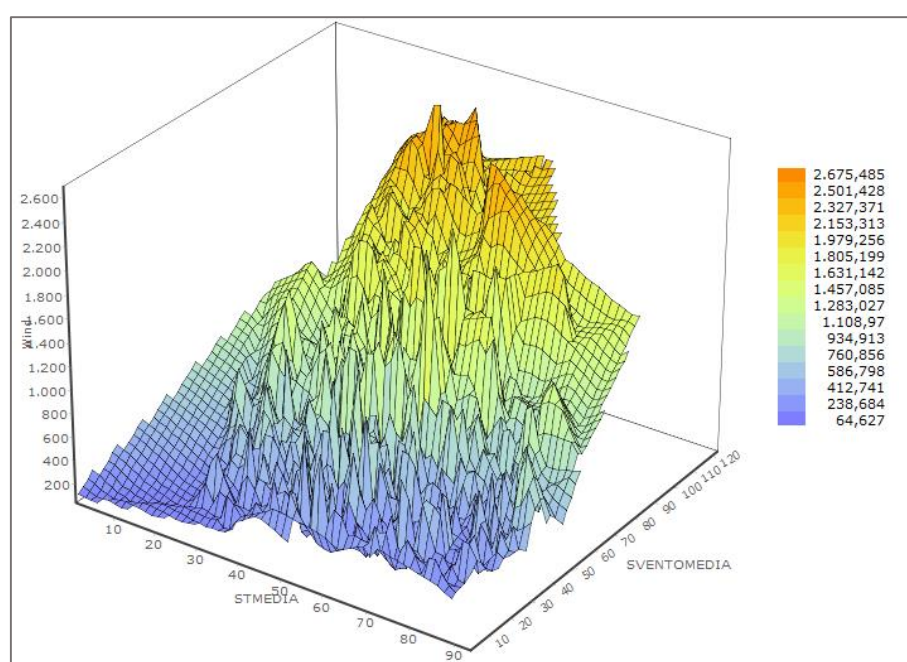


Figura 11.24 - Correlazione tra Eolico tep ed indici climatici medi.



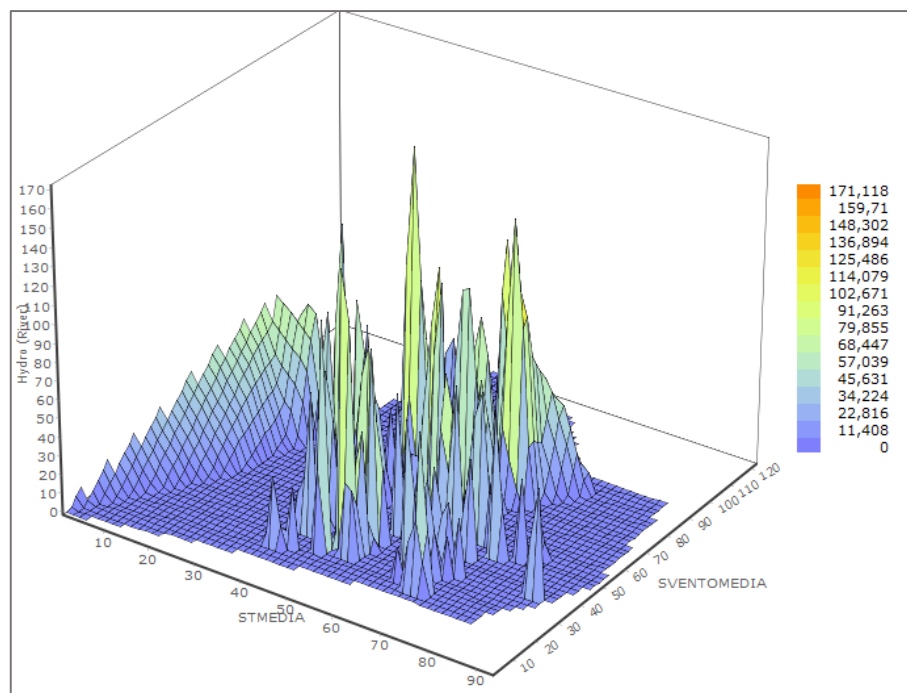


Figura 11.25 - Correlazione tra Idro fl tep ed indici climatici medi.

## 11.6 Modello di Previsione Neural Network “Produzione EE da FER”

Il modello sviluppato con il software GMDH Shell basato sulle Neural Network GMDH ha come scopo la definizione dell’energia prodotta da impianti fotovoltaici, eolici ed idroelettrici fluenti, ponendo come input della rete neurale i dati relativi agli indici ambientali: temperatura, vento, umidità e visibilità; ed i dati relativi alle potenze installate di impianti favolatici ed eolici e di impianti FER nel complesso. Nella Figura 11.26 si riporto lo schema della rete utilizzata.

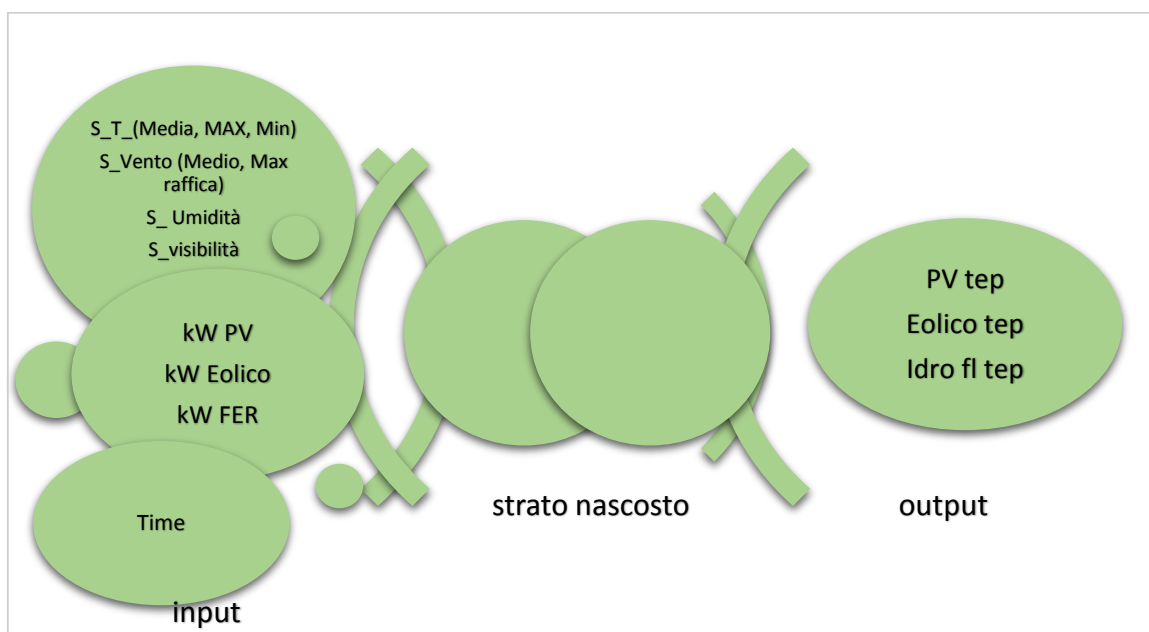


Figura 11.26 – Diagramma neurale - Previsione “Produzione EE da FER”.

Attraverso l'iterazione di più settaggi si è scelto quello che coniugasse basso errore e velocità computazionale. Sono state quindi definite le seguenti impostazioni nel pannello Solver (Tabella 12.7).

|                       |                       |
|-----------------------|-----------------------|
| Reorder observations  | Odd/even              |
| Validation strategy   | k-fold validation     |
| Number of folds       | 2                     |
| Validation criterion  | RMSE                  |
| Variables ranking     | No                    |
| Core algorithm        | GMDH neural network   |
| Neuron function       | $a+x_i+x_i*x_j+x_i^2$ |
| Max. number of layers | 50                    |
| Initial layer width   | 50                    |
| Time series mode      | ✓                     |

Tabella 11.8 – Settings solver – Modello di previsione “Produzione EE da FER”.

In particolare evidenziamo il criterio di validazione RMSE, che rappresenta la deviazione standard del campione delle differenze tra i valori previsti ed i valori osservati, il core algorithm (GMDH neural network) che rappresenta l'algoritmo di apprendimento statistico, la funzione neurale di attivazione del singolo neurone ( $a+x_i+x_i*x_j+x_i^2$ ) con la quale si ottiene il plot del model fit e della predictions, il massimo numero di strati (50) come limite superiore per il numero di strati di rete creati dall'algoritmo, e la larghezza dello strato iniziale (50) che definisce quanti neuroni vengono aggiunti alla serie di ingresso ad ogni nuovo livello.

Avviando il processo di training e di validation si sono quindi ottenuti i seguenti risultati per i singoli target (Tabelle 11.9, 11.10 e 11.11).

| Postprocessed results                     | Model fit | Predictions |
|---|-----------|-------------|
| Number of observations                    | 1169      | 292         |
| Max. negative error                       | -257,74   | -206,76     |
| Max. positive error                       | 396,60    | 273,63      |
| Normalized Mean absolute error (nMAE)     | 8,91 %    | 9,30 %      |
| Normalized Root mean square error (nRMSE) | 11,74 %   | 12,16 %     |
| Residual sum                              | 1,35 e-10 | 1300,93     |
| Standard deviation of residuals           | 80,32     | 83,12       |
| Coefficient of determination (R2)         | 0,76      | 0,76        |
| Correlation                               | 0,877     | 0,874       |

Tabella 11.9 – Accuracy PV tep– Modello di previsione “Produzione EE da FER” con Neural Network.

| Postprocessed results                     | Model fit             | Predictions |
|---|-----------------------|-------------|
| Number of observations                    | 1169                  | 292         |
| Max. negative error                       | -1125,52              | -1419,46    |
| Max. positive error                       | 781,55                | 1106,36     |
| Normalized Mean absolute error (nMAE)     | 7,03 %                | 7,92 %      |
| Normalized Root mean square error (nRMSE) | 9,16 %                | 11,21 %     |
| Residual sum                              | $3,98 \cdot 10^{-10}$ | 5984,87     |
| Standard deviation of residuals           | 258,78                | 315,86      |
| Coefficient of determination ( $R^2$ )    | 0,77                  | 0,71        |
| Correlation                               | 0,878                 | 0,844       |

Tabella 11.10 – Accuracy Eolico tep– Modello di previsione “Produzione EE da FER” con Neural Network.

| Postprocessed results                     | Model fit              | Predictions |
|---|------------------------|-------------|
| Number of observations                    | 1169                   | 292         |
| Max. negative error                       | -191,52                | -117,70     |
| Max. positive error                       | 64,31                  | 167,87      |
| Normalized Mean absolute error (nMAE)     | 1,53 %                 | 15,94 %     |
| Normalized Root mean square error (nRMSE) | 6,22 %                 | 23,21 %     |
| Residual sum                              | $-4,37 \cdot 10^{-10}$ | 7974,76     |
| Standard deviation of residuals           | 12,06                  | 315,86      |
| Coefficient of determination ( $R^2$ )    | 0,80                   | -1,55       |
| Correlation                               | 0,89                   | 0,58        |

Tabella 11.11 – Accuracy Idro fl tep– Modello di previsione “Produzione EE da FER” con Neural Network.

Il calcolo del MAE e del RMSE in forma normalizzata permette di superare l’errore che si genera con il MAPE ed il RMSPE per via della prossimità al valore 0 dei dati di target.

Le figure 11.27, 11.28 ed 11.29 rappresentano il plot del model fit e della previsione. In grigio sono rappresentati i dati reali del DataSet, in blu il risultato del model fit, in rosso il risultato della previsione e la banda rosa rappresenta la banda di confidenza della previsione.

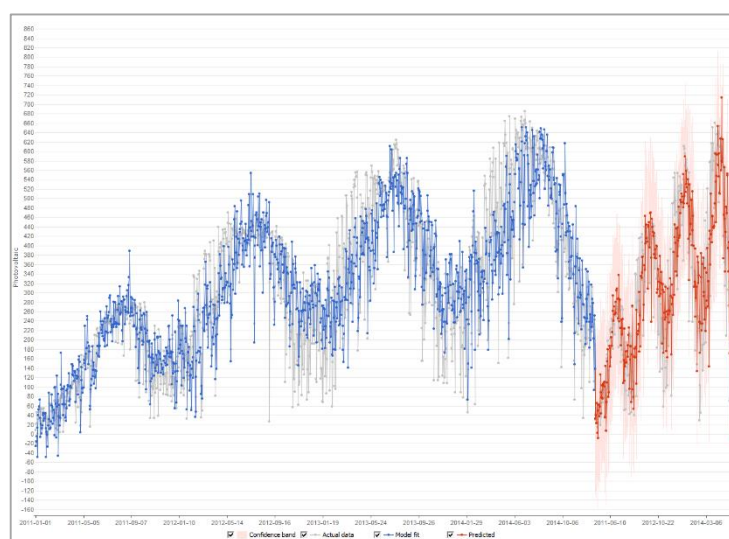


Figura 11.27 – Plot model fit e previsione PV tep– Modello di previsione “Produzione EE da FER” con Neural Network.

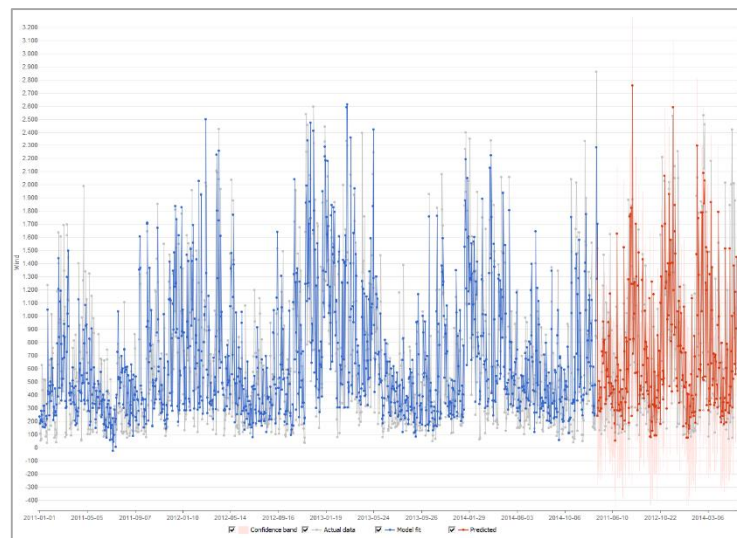


Figura 11.28 – Plot model fit e previsione Eolico tep– Modello di previsione “Produzione EE da FER” con Neural Network.

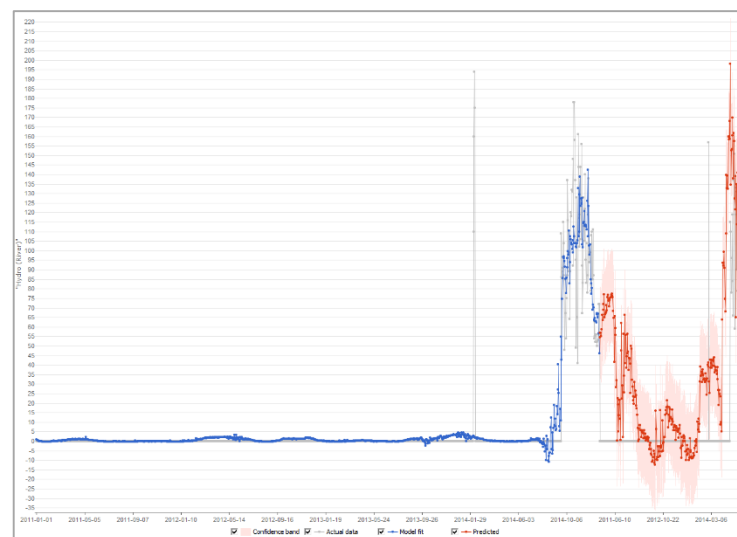


Figura 11.29 – Plot model fit e previsione Idro fl tep– Modello di previsione “Produzione EE da FER” con Neural Network.

Nelle figure 12.30 – 11.38 sono mostrati rispettivamente il “residuals” del model fit e della predictions, l’autocorrelazione e la frequenza del valore residuo.

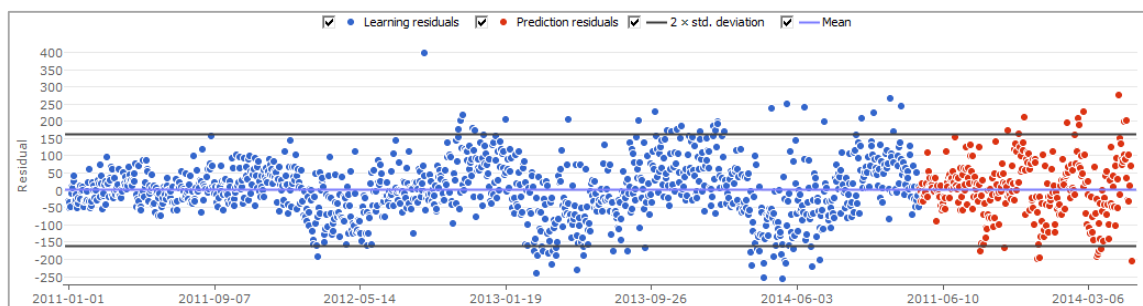


Figura 11.30 – Plot residui previsione PV tep– Modello di previsione “Produzione EE da FER” con Neural Network.

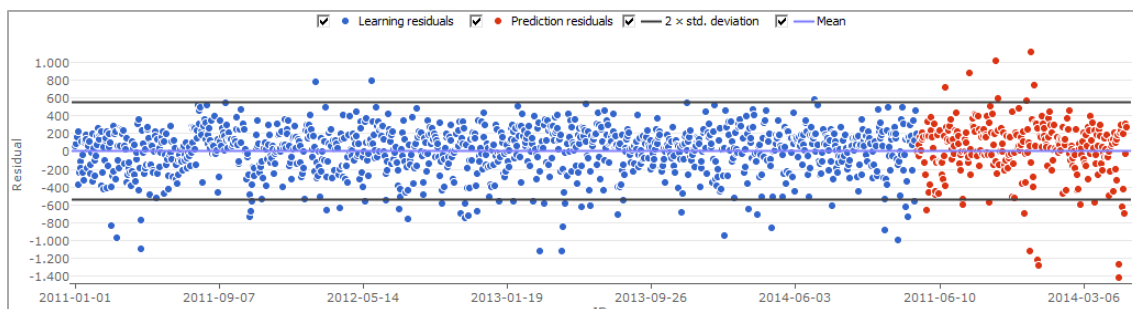


Figura 11.31 – Plot residui previsione Eolico tep– Modello di previsione “Produzione EE da FER” con Neural Network.

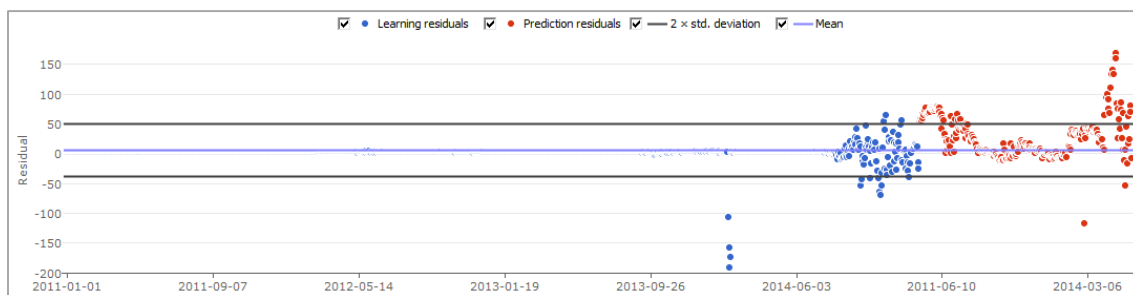


Figura 11.32 – Plot residui previsione Idro fl tep– Modello di previsione “Produzione EE da FER” con Neural Network.

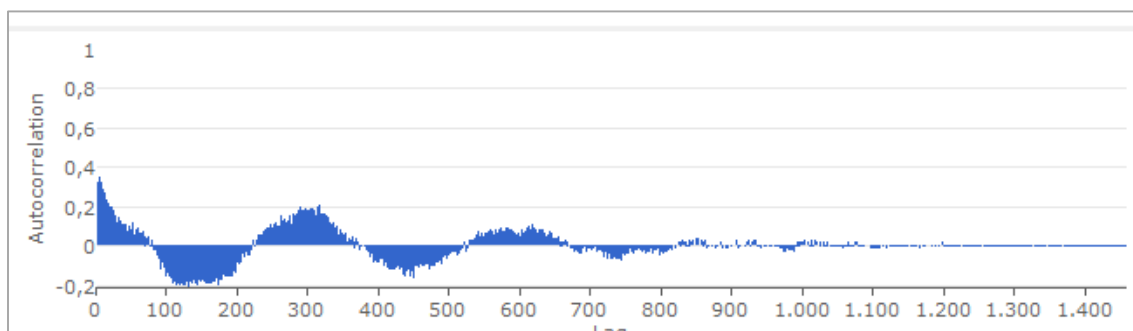


Figura 11.33 – Autocorrelazione previsione PV tep– Modello di previsione “Produzione EE da FER” con Neural Network.

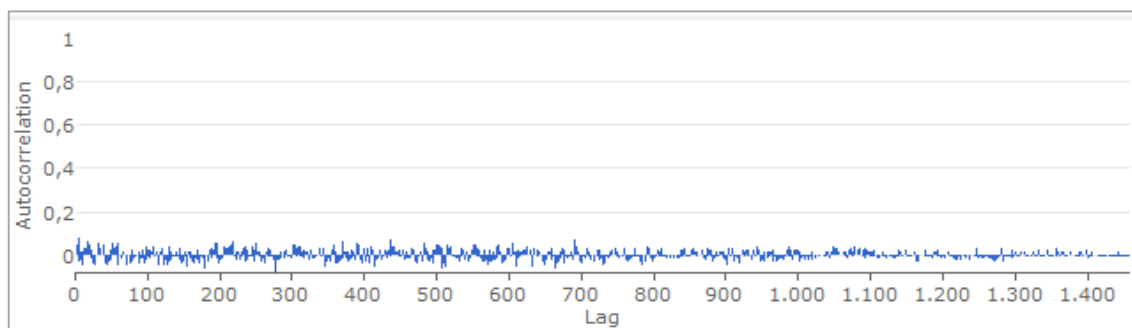


Figura 11.34 – Autocorrelazione previsione Eolico tep– Modello di previsione “Produzione EE da FER” con Neural Network.

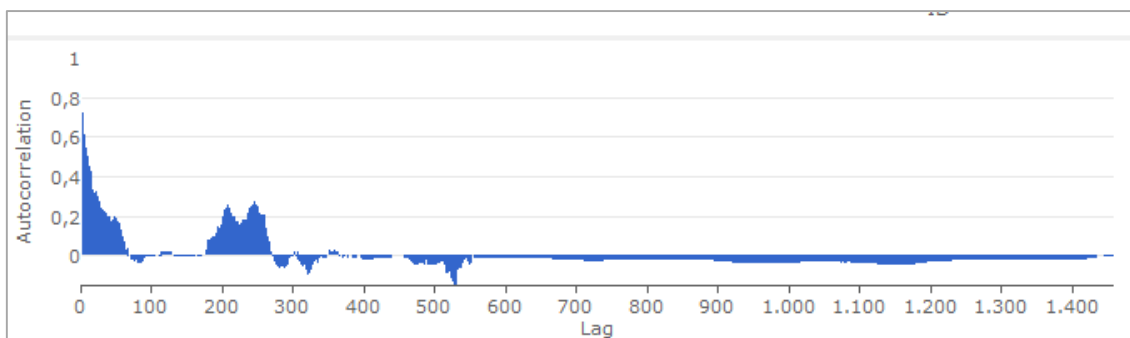


Figura 11.35 – Autocorrelazione previsione Idro fl tep– Modello di previsione “Produzione EE da FER” con Neural Network.

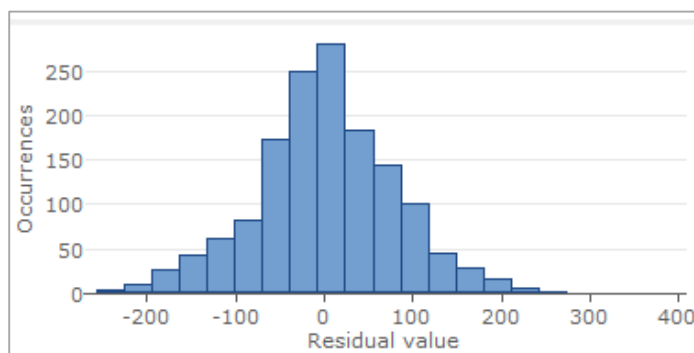


Figura 11.36 – Frequenza residui previsione PV tep– Modello di previsione “Produzione EE da FER” con Neural Network.

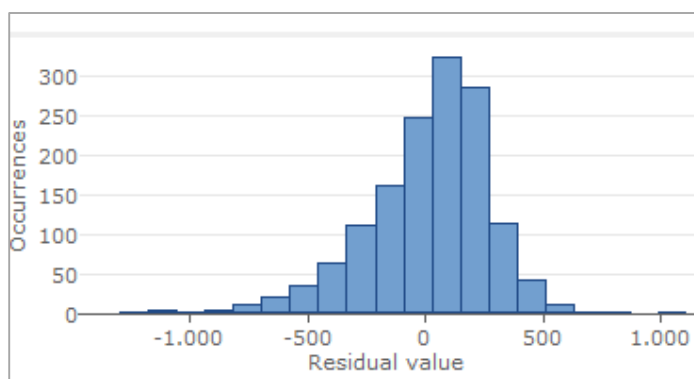


Figura 11.37 – Frequenza residui previsione Eolico tep– Modello di previsione “Produzione EE da FER” con Neural Network.

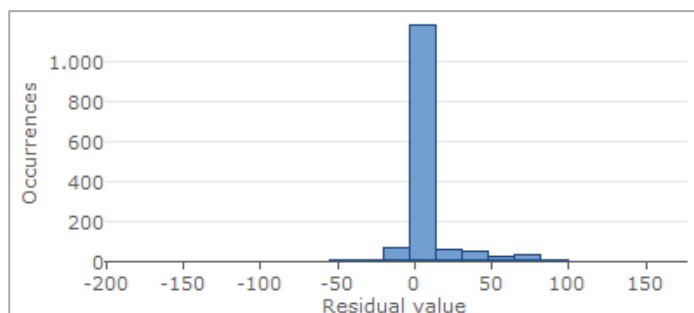


Figura 11.38 – Frequenza residui previsione Idro fl tep– Modello di previsione “Produzione EE da FER” con Neural Network.

Di seguito si riporta le funzioni generate per la predizione.

### PV tep

$$Y1 = -0.0850317 + N5 * 0.58171 + N5^2 * 7.11721e-05 + N9 * 0.391511$$

$N9 = -262.071 + "STMIN, cubert"*155.379 - "STMIN, cubert"*N15*0.342783 - "STMIN, cubert"*20.7458 + N15*1.7673 + N15^2*0.00101364$   
 $N15 = 468.883 + "SUMIDITA, cubert"*77.6397 + "SUMIDITA, cubert"*N25*0.328139 - "SUMIDITA, cubert"*25.1425 - N25*1.05885 + N25^2*6.04912e-05$   
 $N25 = -89.4851 + N41*0.337531 + N41*N64*0.00188645 + N64*0.932393 - N64^2*0.00179932$   
 $N64 = 167.695 - "SVENTOMEDIA, cubert"*78.4845 - "SVENTOMEDIA, cubert"*"SRAFFICA, cubert"*14.5851 + "SVENTOMEDIA, cubert"*41.5118$   
 $N5 = 4.68253 + N55*N12*0.00777546 - N55^2*0.00469987 + N12*1.15775 - N12^2*0.00345828$   
 $N12 = 13.0687 + N28*0.733072 + N28*N51*0.00129597 - N28^2*0.000627179$   
 $N51 = 47.2934 + N58*0.241768 + N58*N60*0.00205587 - N60*0.13441 + N60^2*0.000216789$   
 $N60 = 271.989 - "SVISIBILITA, cubert"*1180.41 + "SVISIBILITA, cubert"*"SVENTOMAX, cubert"*238.709 + "SVISIBILITA, cubert"*75.0905 + "SVENTOMAX, cubert"*622.624 - "SVENTOMAX, cubert"*2*155.246$   
 $N58 = 49443.1 + "SRAFFICA, cubert"*342.767 - "SRAFFICA, cubert"*"Eolico_KW, cubert"*3.08748 - "Eolico_KW, cubert"*890.455 + "Eolico_KW, cubert"*4.01375$   
 $N28 = 612.729 + "SUMIDITA, cubert"*52.8702 + "SUMIDITA, cubert"*N41*0.349499 - "SUMIDITA, cubert"*24.2621 - N41*1.41 + N41^2*0.000439416$   
 $N41 = -425549 - "STMAX, cubert"*3782.41 + "STMAX, cubert"*date*0.0873426 + "STMAX, cubert"*48.2378 + date*20.7729 - date^2*0.000253329$   
 $N55 = 1002.31 - "STMIN, cubert"*299.896 + "STMIN, cubert"*"PV_kW, cubert"*1.51161 + "STMIN, cubert"*42.9675 - "PV_kW, cubert"*19.0454 + "PV_kW, cubert"*2*0.120757$

### Eolico tep

$$Y1 = -16.9595 + N27 * 0.205308 - N27 * N10 * 0.00491633 + N27^2 * 0.00250118 + N10 * 0.765739 + N10^2 * 0.00242712$$

$N10 = 453.224 - N57*1.27576 + N57^2*0.000692717 + N24*1.16445 - N24^2*5.62454e-05$   
 $N24 = 112.913 + N30*0.785105 + N30*N52*0.000483079 - N30^2*7.67973e-05 - N52*0.566901 + N52^2*0.000504557$   
 $N52 = 533.133 + "SVISIBILITA, cubert"*838.564 - "SVISIBILITA, cubert"*235.659 + cycle*0.00698505 - cycle^2*4.23434e-06$   
 $N30 = 395.746 + "STMAX, cubert"*996.955 - "STMAX, cubert"*2*163.459 - "SVENTOMEDIA, cubert"*1500.65 + "SVENTOMEDIA, cubert"*343.396$   
 $N57 = 7.79625e+06 - "Tot_GW, cubert"*264858 + "Tot_GW, cubert"*date*6.89846 - "Tot_GW, cubert"*474.423 - date*280.644 + date^2*0.00211268$   
 $N27 = -101.194 + N31*0.94709 - N31*N45*0.000265142 + N31^2*6.12394e-05 + N45*0.262605 + N45^2*0.000130437$   
 $N45 = 81999.6 - "SRAFFICA, cubert"*90.1157 + "SRAFFICA, cubert"*2*59.816 - "Eolico_KW, cubert"*1413.76 + "Eolico_KW, cubert"*6.11808$   
 $N31 = 384.825 + "STMEDIA, cubert"*1148.68 - "STMEDIA, cubert"*2*191.04 - "SVENTOMEDIA, cubert"*1652.41 + "SVENTOMEDIA, cubert"*2*366.841$

### Idro fl tep

$$Y1 = 0.0128989 + N30 * 0.106029 - N30 * N7 * 0.00982578 + N7 * 0.742278 + N7^2 * 0.0110801$$

$N7 = 1.00887 + N41*0.21097 + N41*N17*0.0451422 - N41^2*0.0122271 - N17^2*0.0199413$   
 $N17 = 1.71076 + N40*0.272724 + N40*N31*0.0583038 - N40^2*0.0301911 + N31*0.607405 - N31^2*0.0251827$   
 $N31 = -1.99669 - N37*0.270722 + N37*N47*0.015111 + N37^2*0.0156113 + N47*0.343906 - N47^2*0.0121973$   
 $N47 = -1.83767 + "STMEDIA, cubert"*138.31 - "STMEDIA, cubert"*2*19.0751 - "SVISIBILITA, cubert"*168.434 + "SVISIBILITA, cubert"*2*29.297$   
 $N37 = -9.73756 + "STMEDIA, cubert"*4.54222 - "STMEDIA, cubert"*cycle*0.209609 - "STMEDIA, cubert"*2*0.436583 + cycle^2*0.0212313$   
 $N40 = 0.849548 + "SRAFFICA, cubert"*cycle*0.0134499 + "SRAFFICA, cubert"*2*0.0149977 - cycle*0.775567 + cycle^2*0.020666$   
 $N30 = -2.05989 + "STMEDIA, cubert"*3.30073 - "STMEDIA, cubert"*N41*0.39966 - "STMEDIA, cubert"*2*0.755307 + N41*1.31786 + N41^2*0.0172312$   
 $N41 = 5.86349 - "SVENTOMAX, cubert"*2.95046 + "SVENTOMAX, cubert"*2*0.412281 - cycle*0.758038 + cycle^2*0.0207005$

## 11.7 Modello di Previsione serie storiche “Produzione EE da FER”

Il modello sviluppato con il software GMDH Shell basato su *Stepwise mixed selection* di GMDH ha come scopo la definizione della produzione di energia elettrica oraria da FER ponendo come input i dati storici orari della variabile di output dall'anno 2011 all'anno 2014.

Dopo diverse iterazioni con più settaggi si è scelto quello che coniugasse basso errore e velocità computazionale. Sono state quindi definite le seguenti impostazioni nel pannello Solver (Tabella 11.12).

|                           |                          |
|---------------------------|--------------------------|
| Reorder observations      | Pseud-random             |
| Validation strategy       | k-fold validation        |
| Number of folds           | 2                        |
| Validation criterion      | RMSE balance             |
| Variables ranking         | By correlation           |
| Drop variables after rank | 25                       |
| Core algorithm            | Stepwise mixed selection |
| Neuron function           | $x_i * x_j$              |
| Limit model complexitt to | 100                      |
| Return set of best models | 50                       |
| Time series mode          | ✓                        |

Tabella 11.12 – Settings solver – Modello di previsione “Produzione EE da FER”.

Avviando il processo di traning con la previsione dei successivi 90 valori, si sono ottenuti i seguenti risultati (Tabelle 11.13, 11.14 ed 11.15).

| Postprocessed results                     | Model fit |
|---|-----------|
| Number of observations                    | 1360      |
| Max. negative error                       | -189,49   |
| Max. positive error                       | 384,20    |
| Normalized Mean absolute error (nMAE)     | 9,36 %    |
| Normalized Root mean square error (nRMSE) | 12,06 %   |
| Residual sum                              | -457,50   |
| Standard deviation of residuals           | 82,54     |
| Coefficient of determination ( $R^2$ )    | 0,72      |
| Correlation                               | 0,85      |

Tabella 11.13 – Accuracy PV tep– Modello di previsione “Produzione EE da FER”.

| Postprocessed results                     | Model fit |
|---|-----------|
| Number of observations                    | 1347      |
| Max. negative error                       | -1869,14  |
| Max. positive error                       | 830,25    |
| Normalized Mean absolute error (nMAE)     | 13,73 %   |
| Normalized Root mean square error (nRMSE) | 17,23 %   |
| Residual sum                              | 19627,1   |
| Standard deviation of residuals           | 486,09    |
| Coefficient of determination ( $R^2$ )    | 0,24      |
| Correlation                               | 0,57      |

Tabella 11.14 – Accuracy Eolico tep– Modello di previsione “Produzione EE da FER”.



| Postprocessed results                     | Model fit |
|---|-----------|
| Number of observations                    | 136       |
| Max. negative error                       | -67,58    |
| Max. positive error                       | 52,86     |
| Normalized Mean absolute error (nMAE)     | 6,67 %    |
| Normalized Root mean square error (nRMSE) | 8,83 %    |
| Residual sum                              | 97,29     |
| Standard deviation of residuals           | 21,87     |
| Coefficient of determination ( $R^2$ )    | 081       |
| Correlation                               | 0,90      |

Tabella 11.15 – Accuracy Idro fl tep– Modello di previsione “Produzione EE da FER”

Le figure 11.39 – 11.41 rappresentano il plot del model fit e della previsione dei singoli target.

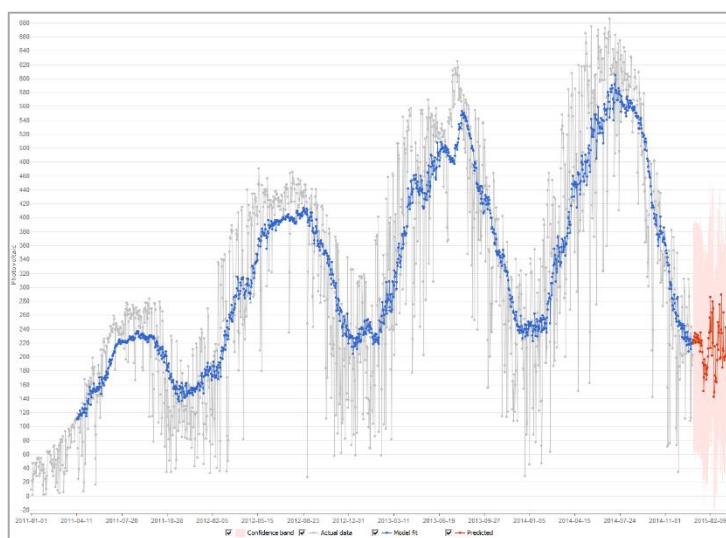


Figura 11.39 – Plot model fit e previsione PV tep “Produzione EE da FER”.

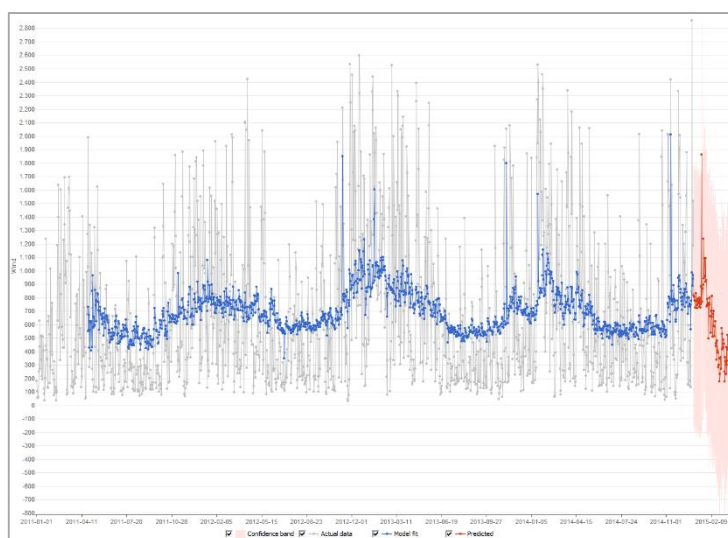


Figura 11.40 – Plot model fit e previsione eolico tep “Produzione EE da FER”.



Figura 11.41 – Plot model fit e previsione Idro fl tep “Produzione EE da FER”.

## 11.8 Conclusioni

Dal confronto dei due modelli implementati si nota come il modello NN rispetto al modello serie storiche dia migliori indici di bontà per i target PV tep ed Eolico tep. Per il target Idro fl tep, invece, il modello serie storiche fornisce indici migliori. Ciò è dovuto dalla pochezza di dati presenti per questo target e quindi da una difficoltà nell'applicazione di modelli NN.

La Tabella 11.14 riporta i valori degli indici di correlazione, i coefficienti  $R^2$  ed i relativi nMAE ed nRMSE.

|   | NN      |          |         | Serie storiche |          |         |
|---|---------|----------|---------|----------------|----------|---------|
|   | PV      | Eolico   | Idro fl | PV             | Eolico   | Idro fl |
| Number of observations                    | 292     | 292      | 292     | 1360           | 1347     | 136     |
| Max. negative error                       | -206,76 | -1419,46 | -117,70 | -189,49        | -1869,14 | -67,58  |
| Max. positive error                       | 273,63  | 1106,36  | 167,87  | 384,20         | 830,25   | 52,86   |
| Normalized Mean absolute error (nMAE)     | 9,30 %  | 7,92 %   | 15,94 % | 9,36 %         | 13,73 %  | 6,67 %  |
| Normalized Root mean square error (nRMSE) | 12,16 % | 11,21 %  | 23,21 % | 12,06 %        | 17,23 %  | 8,83 %  |
| Residual sum                              | 1300,93 | 5984,87  | 7974,76 | -457,50        | 19627,1  | 97,29   |
| Standard deviation of residuals           | 83,12   | 315,86   | 315,86  | 82,54          | 486,09   | 21,87   |
| Coefficient of determination ( $R^2$ )    | 0,76    | 0,71     | -1,55   | 0,72           | 0,24     | 0,81    |
| Correlation                               | 0,874   | 0,844    | 0,58    | 0,85           | 0,57     | 0,90    |

Tabella 11.14 – Confronto modelli di previsione “Produzione EE da FER”

La quasi uguaglianza dei due modelli in termini di indicatori di bontà, tranne per il caso del idro fl tep, è essenzialmente dovuta dalla buona organizzazione del DataSet, (serie continue e pluriennali).

E' quindi evidente che la scelta degli indicatori ambientali (temperatura, umidità, visibilità e vento), di impianto (potenza installata) e temporali (periodo dell'anno) rendono il modello predittivo con Neural Network molto performante ed in grado di predire sia la stagionalità ma anche le variazioni giornaliere della produzione da impianti eolici e fotovoltaici. Per quanto riguarda la produzione di energia da idroelettrico fluente, data la natura stagionale della fonte di produzione (fenomeni meteorologici) è più conveniente effettuare la predizione con modelli basati su serie storiche.

## CAPITOLO 12 *Caso studio*

# *Prezzo dell'energia elettrica*

Grazie al decreto “Bersani” di liberalizzazione il settore elettrico è profondamente cambiato. Nel 1999 le condizioni erano molto critiche; la situazione infrastrutturale nel settore elettrico era non solo al limite del collasso ma “ingessata” dalla difficoltà di realizzare nuove centrali e nuove linee di trasmissione; non esisteva un mercato ma solo grandi e piccoli monopoli; i prezzi erano ben superiori a quelli degli altri Paesi europei nonostante un prezzo del petrolio che viaggiava intorno ai 10-20 \$/barile; non esisteva una diffusa consapevolezza da parte dei consumatori sulle tematiche inerenti il costo della energia.

Gli effetti della liberalizzazione non furono, peraltro, immediati; basti ricordare che il 2003 fu un anno orribile per l'energia elettrica: il black out di settembre fu solo il tragico culmine di una lunga serie di criticità.

Oggi l'Italia ha un parco di generazione completamente rinnovato tra i più efficienti del mondo, un mercato competitivo e un gestore indipendente della rete di trasmissione, Terna S.p.A., che investe cinque volte di più rispetto a quando era una costola dell'Enel. I consumatori godono di un servizio migliore in termini di sicurezza, di minori interruzioni e di maggiori diritti verso i propri fornitori; gli utenti sono diventati clienti; milioni di clienti hanno cambiato il loro fornitore e le associazioni dei consumatori sono diventate un riferimento affidabile e competente.

Ma la liberalizzazione ha garantito risultati tangibili anche sotto il profilo economico: l'Autorità per l'energia elettrica e il gas ha valutato che nel settore elettrico del nostro Paese si è ottenuta una riduzione di oneri stimabile in più di 4,5 miliardi di euro all'anno, rispetto al 1999, in termini di minor consumo di combustibili, di minori emissioni di CO<sub>2</sub> e di efficienze gestionali.

I prezzi dell'energia elettrica continuano ad essere superiori rispetto a quelli di molti Paesi europei ma, nonostante il prezzo del petrolio sia cresciuto di almeno 5 volte, è ormai in corso un percorso di progressiva riduzione del differenziale di prezzo con Paesi come la Francia e la Germania dove è ben più ampio il ricorso a fonti diverse dagli idrocarburi.

Ormai da diversi anni l'Italia non è più importatrice di energia elettrica per necessità ma solo per convenienza, ed anzi in particolari periodi di punta è l'Italia che esporta energia verso l'Europa.

Tutto ciò è senza dubbio un buon risultato per il settore elettrico, frutto di una regolazione molto avanzata, di meccanismi tariffari evoluti e della maggiore concorrenza; basti pensare che si è ridotta a

meno del 30% la quota dell'operatore ex monopolista Enel, che ha comunque saputo compensare con una significativa penetrazione nel più vasto mercato internazionale.

Se oggi il settore elettrico potesse approvvigionarsi di gas ai prezzi del mercato USA, i prezzi del mercato elettrico all'ingrosso italiano sarebbero minori di quelli del mercato francese o tedesco: intervenire nell'assetto del sistema del gas è la via maestra per risolvere anche il problema della competitività dell'energia elettrica.

Nei circa dieci anni di funzionamento del Mercato Elettrico sono intervenuti tre cambiamenti rilevanti:

- il primo è stato un forte incremento della potenza termoelettrica installata, che da una parte ha incrementato la sicurezza e la concorrenza ma dall'altra ha concorso a creare una situazione di overcapacity;
- il secondo è stato la realizzazione di numerose nuove linee di trasmissione, che hanno contribuito a ridurre gli oneri di congestione; rimangono tuttavia ancora alcune importanti linee, come quella dalla Calabria alla Sicilia, che devono ancora essere completate;
- il terzo è stato il rapido incremento di impianti da fonte rinnovabile che ha invece causato importanti problemi di funzionamento al mercato elettrico.

### **12.1 I problemi indotti dalle rinnovabili al mercato**

I principali problemi indotti dalle fonti rinnovabili al funzionamento del mercato elettrico sono i seguenti:

- si è ridotta la quota di impianti che operano in regime di concorrenza; infatti le fonti rinnovabili nel mercato godono della priorità di dispacciamento (garantita dalle direttive europee) e quindi agiscono da “price takers”. In sostanza il meccanismo di soluzione del mercato basato sul prezzo marginale (tutti gli impianti vengono remunerati al prezzo dell'impianto più costoso che viene selezionato nel Mercato del Giorno Prima) consente alle fonti rinnovabili di offrire a prezzo zero e di essere remunerate al prezzo dell'impianto termoelettrico più costoso senza partecipare al mercato; la concorrenza avviene quindi solo tra gli impianti termoelettrici;
- l'esito del Mercato del Giorno Prima è sempre meno eseguibile. In sostanza, nei primi anni di funzionamento del MGP l'esito economico del mercato coincideva largamente con il reale funzionamento degli impianti e i mercati successivi avevano la funzione di ottimizzare l'esito del MGP e di consentire la gestione degli eventi che intercorrevano tra la chiusura del mercato e l'effettiva chiamata in esercizio degli impianti; con una quota rilevante di fonti rinnovabili accade invece sempre più spesso che l'esito del MGP non sia eseguibile perché gli impianti selezionati in base al merito economico non garantiscono un livello sufficiente di riserve

(primaria, secondaria e terziaria) che non possono essere fornite dagli impianti fotovoltaici o eolici; i mercati successivi hanno quindi assunto la funzione impropria di “correggere” l’esito del MGP per garantire la sicurezza del sistema. Naturalmente ogni modifica è onerosa;

- si è ridotta la sicurezza del sistema; il fatto che l’esito del MGP non sia eseguibile a causa di un insufficiente livello di riserve (primaria, secondaria e terziaria) può incidere, in particolari situazioni, anche sul livello di sicurezza. Ad esempio quando il livello di produzione degli impianti fotovoltaici ed eolici raggiunge le dimensioni della domanda nella stessa zona (tipicamente in estate con basso livello di domanda, in ore diurne e in giornate ventose) l’esito del MGP diventa pari a zero in termini di prezzo e l’intervento di Terna S.p.A. per garantire la sicurezza del sistema consiste non solo nel “chiamare” impianti termoelettrici nei mercati successivi ma anche nel fermare una quota di impianti rinnovabili. Questa situazione si è verificata, a partire dal 2012, in alcune centinaia di ore e diventerà sempre più frequente con l’aumento della produzione rinnovabile. Tanto maggiore è l’intervento di Terna S.p.A. in prossimità del momento di consegna (ovvero nel momento in cui gli impianti devono eseguire le disposizioni del Gestore) tanto più aumenta la difficoltà di gestire in sicurezza il sistema;
- esiste una mancanza di equità tra la remunerazione degli impianti termoelettrici e la remunerazione delle fonti rinnovabili. Il MGP non distingue e valorizza allo stesso modo l’energia offerta da impianti termoelettrici e quella offerta da impianti rinnovabili non programmabili, che non solo è più incerta e quindi richiede la disponibilità di riserve, ma non può offrire servizi di riserva. Questa iniquità è ciò che genera la richiesta di un “capacity payment” a favore degli impianti termoelettrici motivata dal fatto che è la stessa esistenza di tali impianti a permettere il funzionamento del sistema elettrico.

## **12.2 La base dati**

I dati sono stati estratti prevalentemente da Terna S.p.a. [38] e dal Gestore dei Mercati Energetici (GME) [47] e riguardano la produzione giornaliera degli impianti e gli esiti delle quantità scambiate sul MGP ed i relativi prezzi su base oraria.

I dati raccolti interessano il periodo che va dal 2011 al 2014 per Terna S.p.a. e dal 2004 al 2015 per il GME. Al fine di rendere i dati confrontabili ed utilizzabili contemporaneamente l’analisi ed il modello sono stati sviluppati su un campione di anni 2011-2014 riportando i dati all’unità di misura Tep.

I dati sono organizzati su base oraria per le diverse zone elettriche, quindi sono state valutate ed accorpate le informazioni sui dati di “Acquisti” e “Vendite”, ed il prezzo zonale orario per le zone “Priolo Gargallo” e “Sicilia” che costituiscono la macro area Sicilia.

Successivamente è stato introdotto un parametro definito “Delta” e che rappresenta il differenziale tra le voci di Vendita e Acquisto.

Con il termine “Vendita” intendiamo tutte le offerte degli operatori in borsa e fuori borsa, mentre con il termine “Acquisto” intendiamo il fabbisogno, le richieste per i pompaggi e gli export verso le zone estere.

La Tabella 3.1 che segue è solo un esempio, limitato alle ore del solo giorno 01/01/2011, di come i dati sono stati organizzati. Tale organizzazione deve essere estesa ai 35.065 valori del campione.

| Data     | h  | Acquisti<br>[tep] | Vendite<br>[tep] | Prezzi<br>[€/tep] | Delta<br>[tep] | Idro<br>[tep] | PV<br>[tep] | P.ggi<br>[tep] | Eolico<br>[tep] | Term.co<br>[tep] |
|----------|----|-------------------|------------------|-------------------|----------------|---------------|-------------|----------------|-----------------|------------------|
| 20110101 | 1  | 2,112             | 1,946            | 0,0524            | -0,166         | 0,086         | 0           | -2,064         | 173,118         | 8,084            |
| 20110101 | 2  | 2,017             | 1,954            | 0,05464           | -0,063         | 0,086         | 0           | -2,408         | 166,066         | 5,59             |
| 20110101 | 3  | 1,96              | 1,855            | 0,06712           | -0,105         | 0,086         | 0           | -2,408         | 158,584         | 4,644            |
| 20110101 | 4  | 2,035             | 2,135            | 0,05526           | 0,1            | 0,086         | 0           | -2,408         | 149,898         | 4,73             |
| 20110101 | 5  | 2,066             | 1,802            | 0,055             | -0,264         | 0,086         | 0           | -15,308        | 152,306         | 7,912            |
| 20110101 | 6  | 1,969             | 1,793            | 0,05402           | -0,176         | 0,086         | 0           | -15,308        | 148,092         | 12,298           |
| 20110101 | 7  | 1,896             | 1,792            | 0,05401           | -0,104         | 0,086         | 0           | -12,986        | 149,554         | 15,824           |
| 20110101 | 8  | 1,732             | 1,785            | 0,054             | 0,053          | 0,258         | 0,086       | -12,9          | 152,736         | 18,834           |
| 20110101 | 9  | 1,573             | 1,783            | 0,05402           | 0,21           | 0,43          | 1,118       | -12,9          | 164,26          | 19,866           |
| 20110101 | 10 | 1,548             | 1,798            | 0,015             | 0,25           | 0,43          | 1,29        | -12,814        | 171,914         | 20,726           |
| 20110101 | 11 | 1,698             | 1,806            | 0,054             | 0,108          | 0,43          | 1,72        | -1,204         | 176,73          | 14,792           |
| 20110101 | 12 | 1,909             | 1,813            | 0,06004           | -0,096         | 0,516         | 1,892       | 0              | 190,92          | 12,642           |
| 20110101 | 13 | 2,086             | 1,986            | 0,097             | -0,1           | 0,516         | 1,032       | 0              | 196,51          | 10,578           |
| 20110101 | 14 | 2,211             | 2,111            | 0,09701           | -0,1           | 0,43          | 0,602       | 0              | 179,138         | 7,138            |
| 20110101 | 15 | 2,229             | 2,129            | 0,09701           | -0,1           | 0,43          | 0,688       | 0              | 161,852         | 5,074            |
| 20110101 | 16 | 2,019             | 1,919            | 0,097             | -0,1           | 0,43          | 0,344       | 0              | 163,056         | 3,698            |
| 20110101 | 17 | 1,827             | 1,817            | 0,061             | -0,01          | 0,344         | 0           | 0              | 176,472         | 2,408            |
| 20110101 | 18 | 1,817             | 1,791            | 0,061             | -0,026         | 0,43          | 0           | 0              | 204,68          | 2,064            |
| 20110101 | 19 | 1,908             | 1,838            | 0,069             | -0,07          | 3,612         | 0           | 0              | 206,916         | 1,806            |
| 20110101 | 20 | 2,207             | 2,107            | 0,097             | -0,1           | 8,17          | 0           | 0              | 211,99          | 1,29             |
| 20110101 | 21 | 2,283             | 2,183            | 0,097             | -0,1           | 7,826         | 0           | 0              | 210,012         | 0,688            |
| 20110101 | 22 | 2,314             | 2,214            | 0,09701           | -0,1           | 0,344         | 0           | 0              | 211,474         | 0,602            |
| 20110101 | 23 | 2,342             | 2,242            | 0,09701           | -0,1           | 0,344         | 0           | 0              | 193,328         | 0,344            |
| 20110101 | 24 | 2,278             | 2,178            | 0,09701           | -0,1           | 0,344         | 0           | 0              | 183,008         | 0,43             |

Tabella 12.1 - DataSet.

## 12.3 L'analisi dati

Tramite l'utilizzo del pacchetto IBM® SPSS® Statistics è stata svolta l'analisi descrittiva dei dati del campione e generati i grafici di frequenza. Tale operazione è stata condotta per l'intero campione di 35.064 dati e per le 8760 ore degli anni 2011, 2013 e 2014, e per le 8784 dell'anno 2012.

Nella Tabella 12.2 è possibile quindi prendere visione delle caratteristiche del campione analizzato.

| Statistiche      |       |          |          |         |          |           |           |            |            |           |
|------------------|-------|----------|----------|---------|----------|-----------|-----------|------------|------------|-----------|
|                  | h     | Acquisti | Vendite  | Prezzi  | Delta    | Idro      | PV        | P.ggi      | Term.co    | Eolico    |
| N Validi         | 35064 | 35064    | 35064    | 35064   | 35064    | 35064     | 35064     | 35064      | 35064      | 35064     |
| N Mancanti       | 0     | 0        | 0        | 0       | 0        | 0         | 0         | 0          | 0          | 0         |
| Media            |       | 2,2032   | 2,0877   | ,0903   | -,1155   | 4,2061    | 12,9576   | -4,3714    | 155,2460   | 27,8159   |
| E.S. della media |       | ,00211   | ,00228   | ,00021  | ,00065   | ,03568    | ,10382    | ,04264     | ,18877     | ,13974    |
| Mediana          |       | 2,2040   | 2,1110   | ,0860   | -,1000   | 1,2040    | ,0860     | ,0000      | 151,8760   | 18,9200   |
| Moda             |       | 2,26     | 2,15     | ,06     | -,10     | ,09       | ,00       | ,00        | 151,27     | 1,98      |
| Deviazione std.  |       | ,39593   | ,42771   | ,03864  | ,12214   | 6,68060   | 19,44143  | 7,98457    | 35,34865   | 26,16709  |
| Varianza         |       | ,157     | ,183     | ,001    | ,015     | 44,630    | 377,969   | 63,753     | 1249,527   | 684,716   |
| Intervallo       | 23,00 | 2,31     | 2,52     | ,30     | ,55      | 47,04     | 79,38     | 43,17      | 260,92     | 199,35    |
| Minimo           | 1,00  | 1,14     | ,84      | ,00     | -,30     | ,00       | ,00       | -43,17     | 57,88      | ,00       |
| Massimo          | 24,00 | 3,45     | 3,36     | ,30     | ,25      | 47,04     | 79,38     | ,00        | 318,80     | 199,35    |
| Somma            | 43830 | 77251,87 | 73203,38 | 3164,59 | -4048,49 | 147482,95 | 454346,69 | -153277,97 | 5443546,86 | 975338,08 |

Tabella 12.2 - Statistiche descrittive DataSet.

Di seguito si ripotano i grafici relativi alla frequenza per singolo fattore.

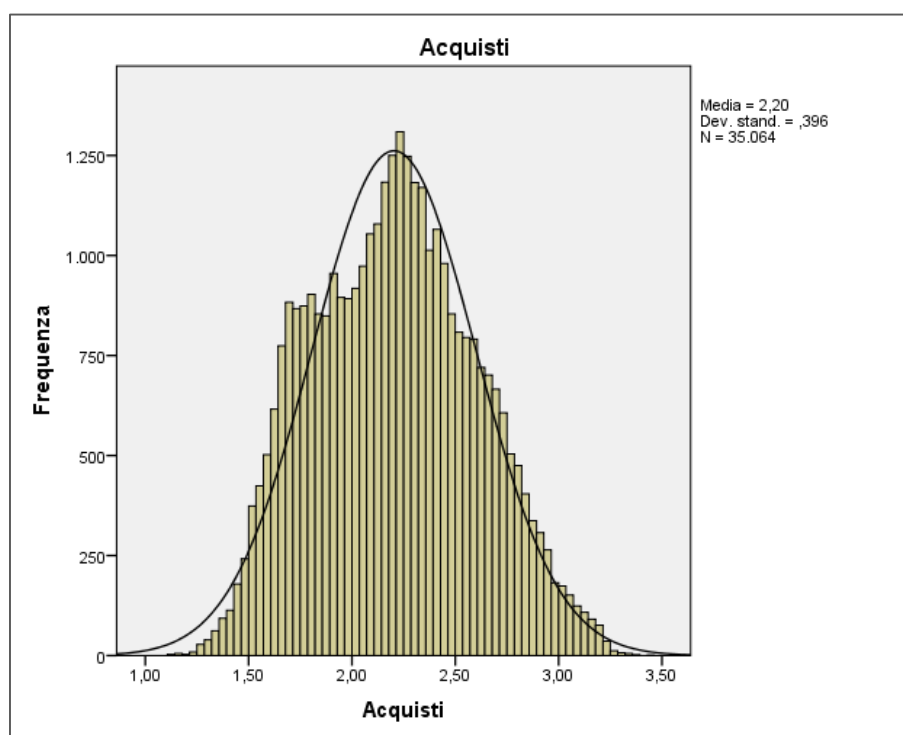


Figura 12.1 - Frequenza valori in tep degli Acquisti dal 2011 al 2014.



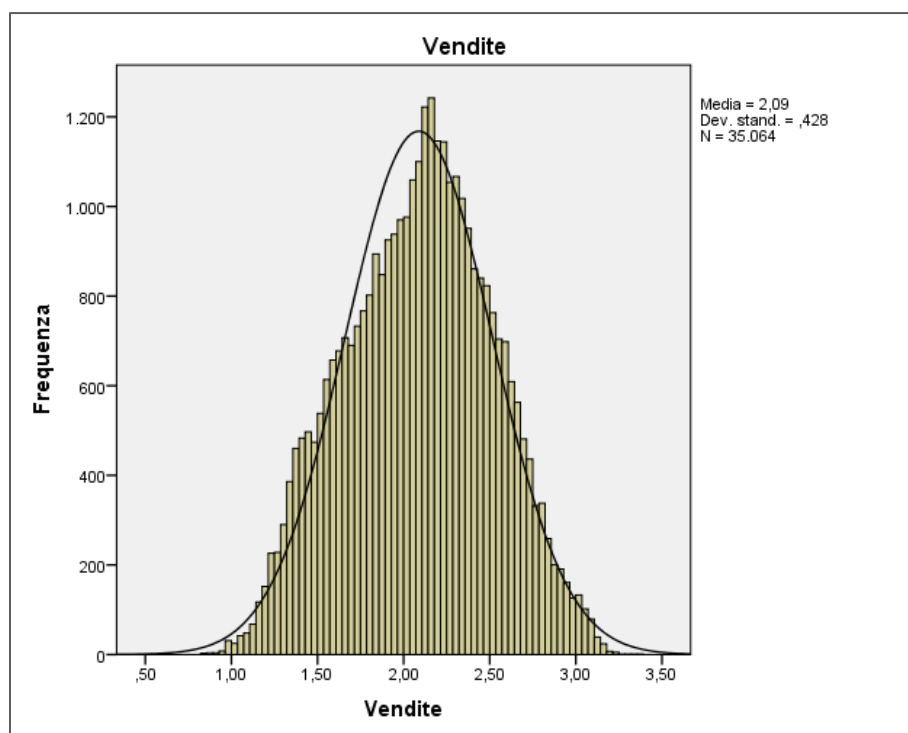


Figura 12.2 - Frequenza valori in tep delle Vendite dal 2011 al 2014.

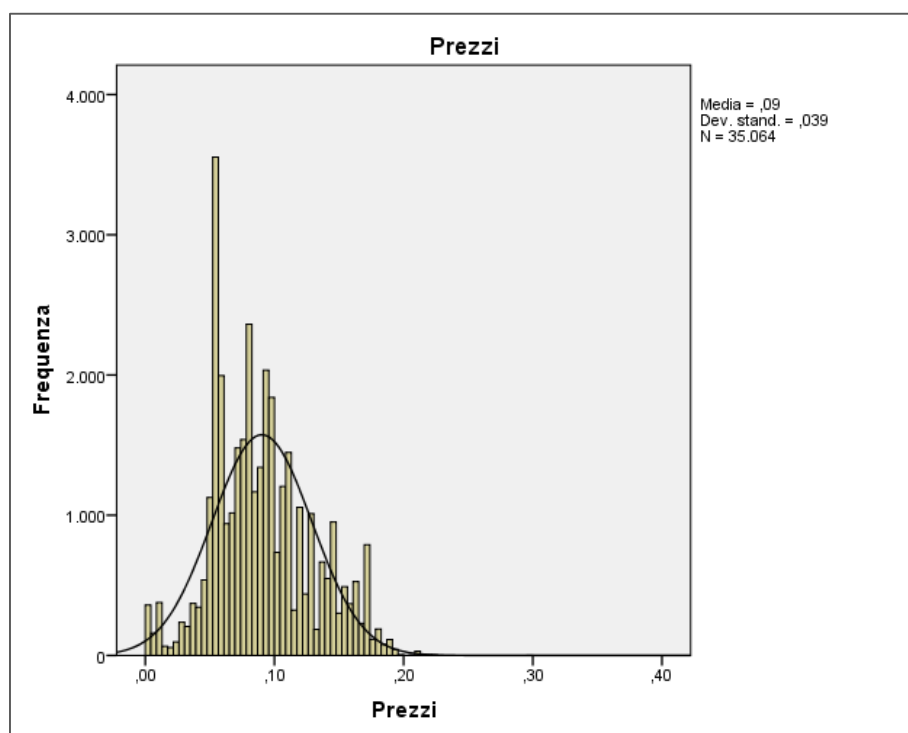


Figura 12.3 - Frequenza valori in €/tep dei Prezzi dal 2011 al 2014.

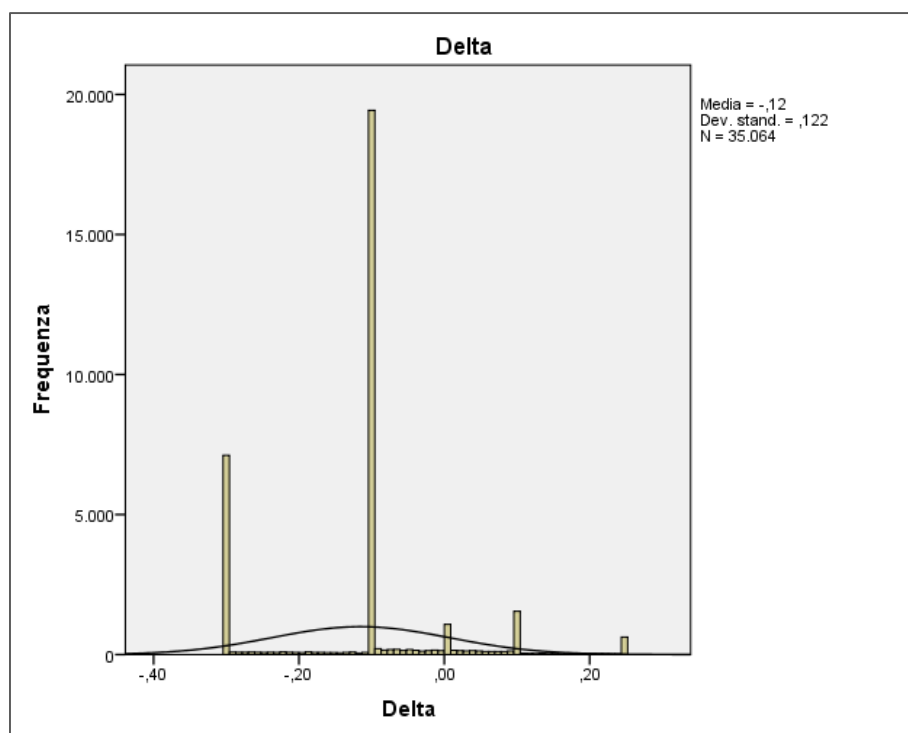


Figura 12.4 - Frequenza valori in tep dei delta dal 2011 al 2014.

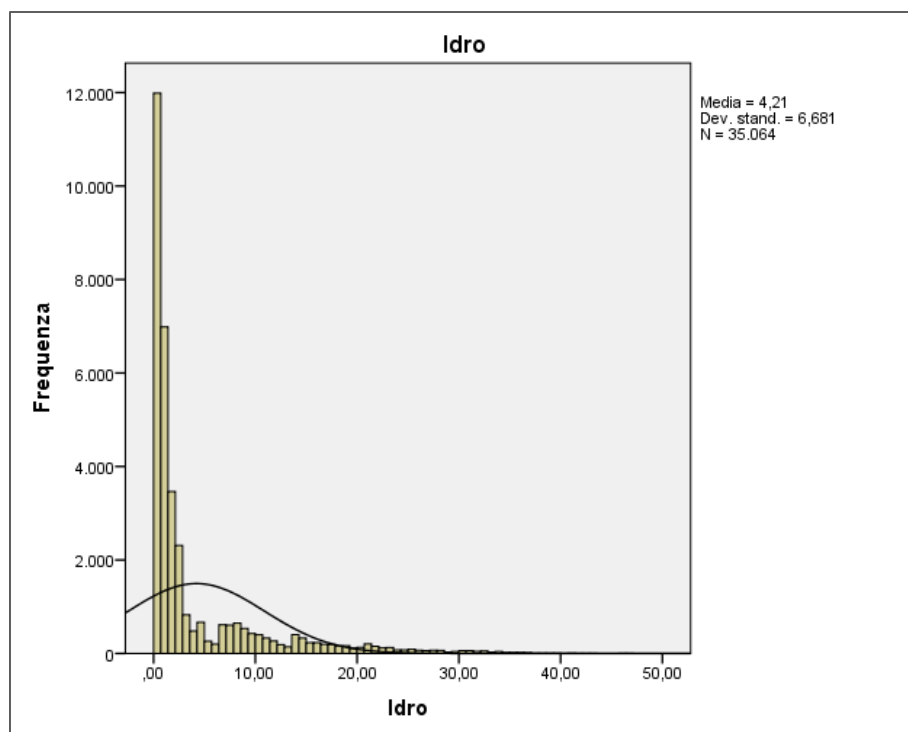


Figura 12.5 - Frequenza valori in tep della produzione oraria da idroelettrico dal 2011 al 2014.

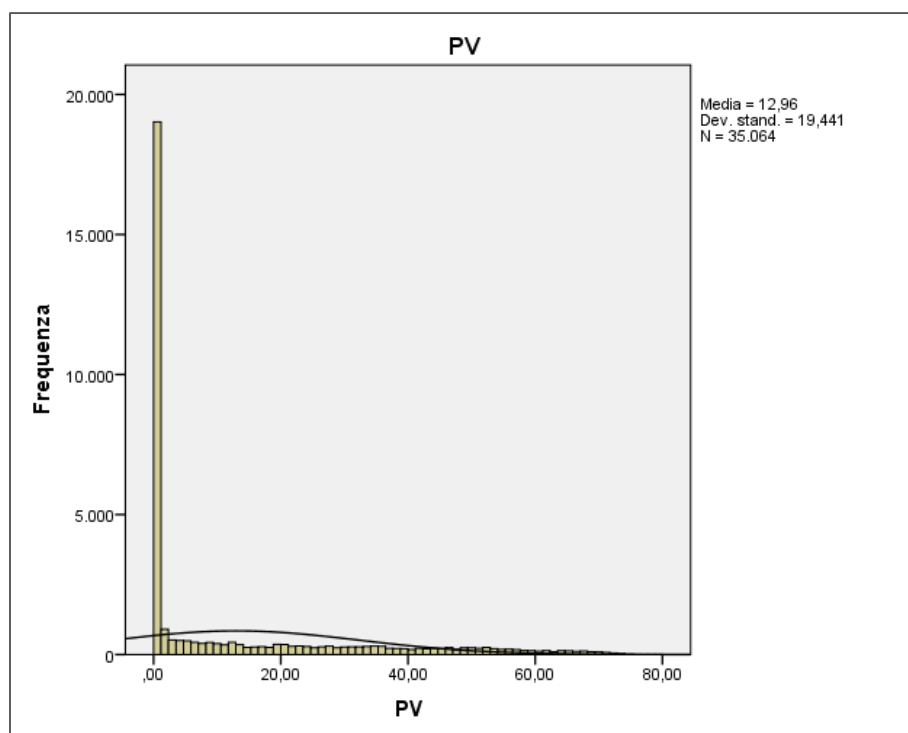


Figura 12.6 - Frequenza valori in tep della produzione oraria da fotovoltaico dal 2011 al 2014.

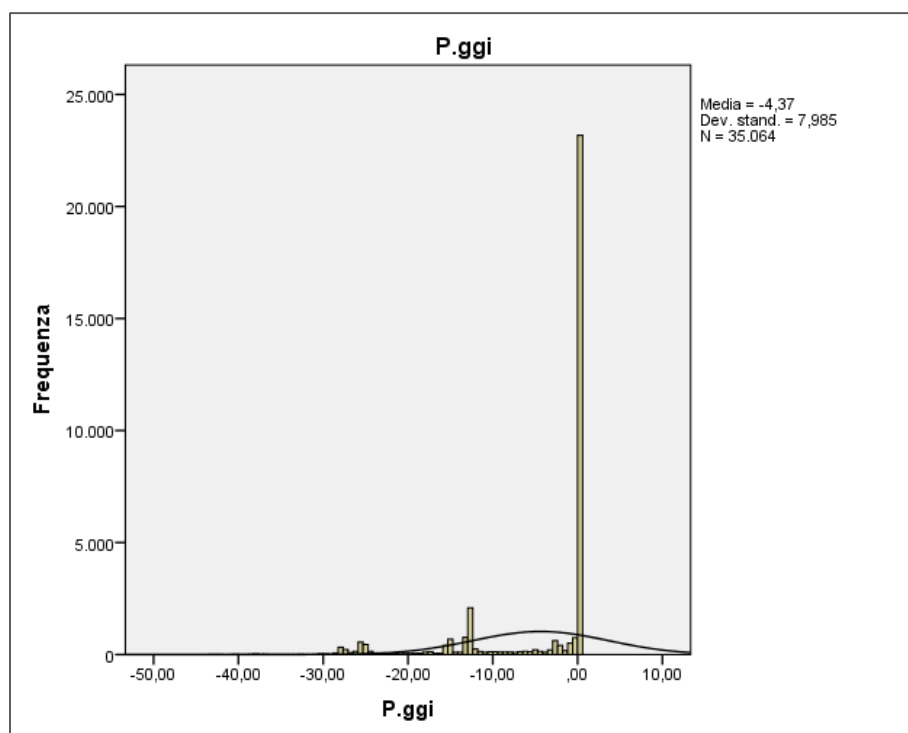


Figura 12.7 - Frequenza valori in tep del consumo orario dei pompaggi dal 2011 al 2014.

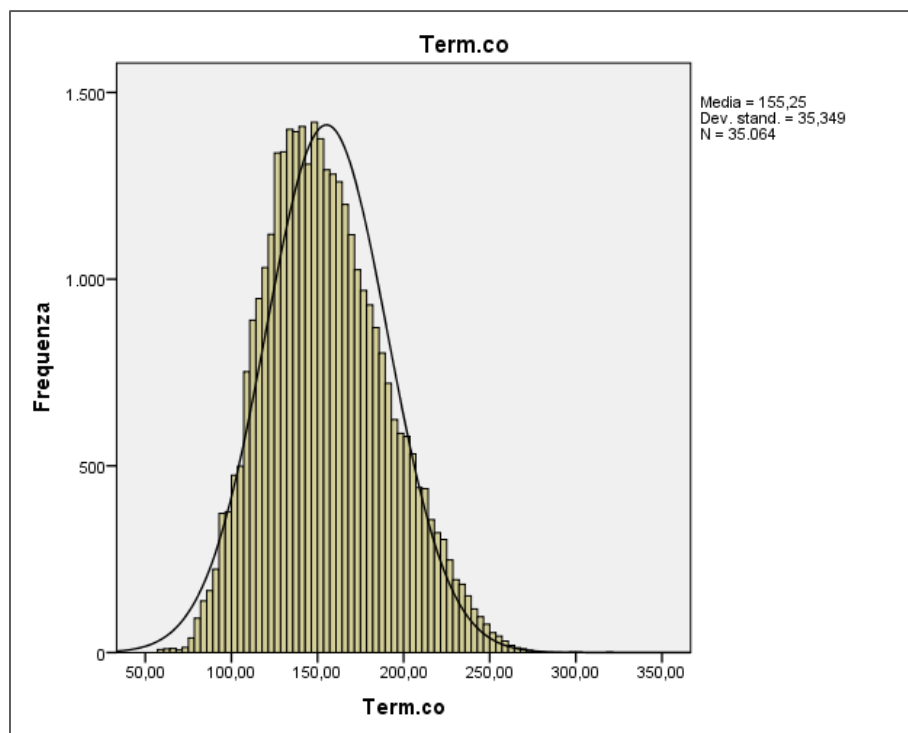


Figura 12.8 - Frequenza valori in tep della produzione oraria da termoelettrico dal 2011 al 2014.

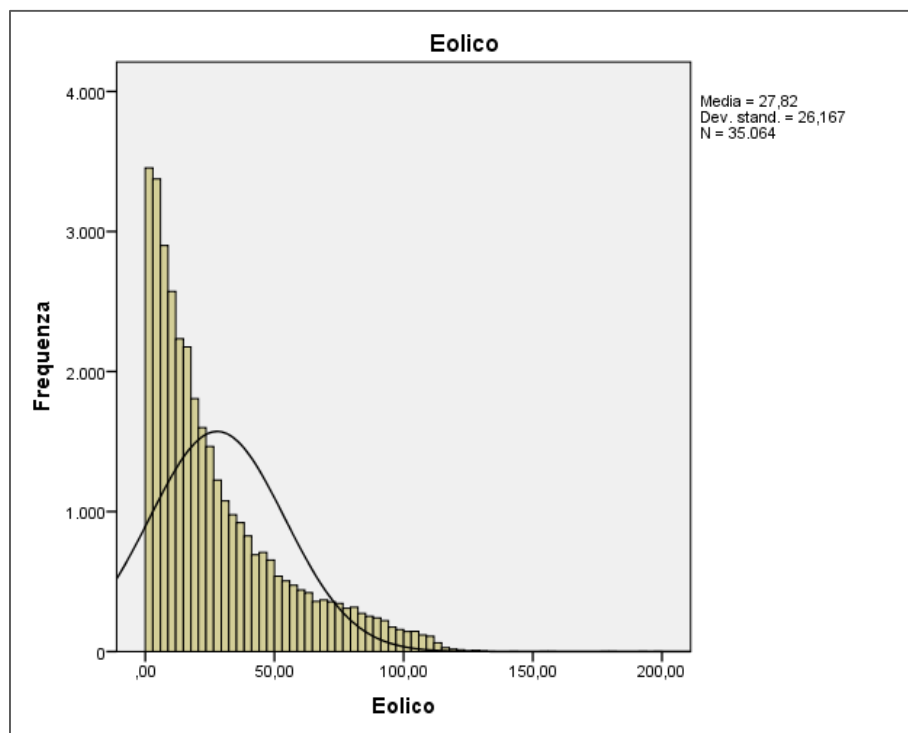


Figura 12.9 - Frequenza valori in tep della produzione oraria da eolico dal 2011 al 2014.

È tuttavia opportuno per apprezzare il cambiamento avuto nei singoli anni, determinato dalla variazione delle potenze installate di impianti alimentati da fonti rinnovabili, comparare i dati dei singoli anni, alla luce delle potenze installate riportate nella Tabella 12.3 e Figura 12.10.

| Tecnologia     | 2011   | 2012   | 2013   | 2014   |
|----------------|--------|--------|--------|--------|
| Idroelettrico  | 720,3  | 720,3  | 719,1  | 722,2  |
| Termoelettrico | 5615,6 | 5642,2 | 5675,7 | 5439,4 |
| Eolico         | 1680,0 | 1749,2 | 1750,2 | 1747,4 |
| Fotovoltaico   | 865,0  | 1125,9 | 1256,2 | 1294,9 |

Tabella 12.3 - Potenza installata in MW per singola fonte dal 2011 al 2014.

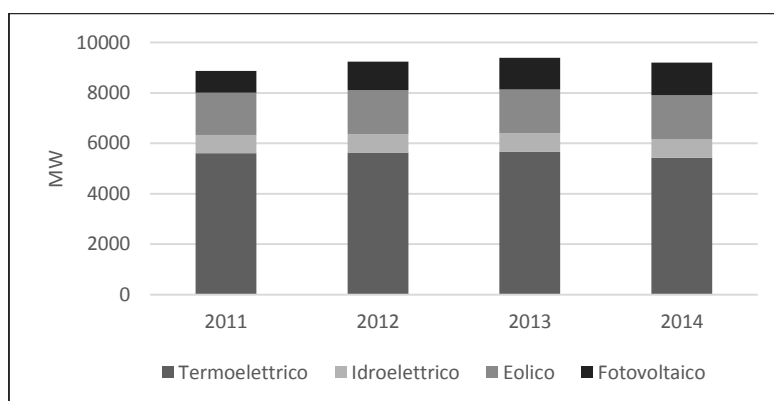


Figura 12.10 - Potenza installata in MW per singola fonte dal 2011 al 2014.

Di seguito si riportano la Tabelle di descrizione dati per singolo anno elaborate tramite il pacchetto IBM® SPSS® Statistics (Tabella 12.3 – 12.6)

| Statistiche      |       |          |          |        |         |          |          |           |            |           |
|------------------|-------|----------|----------|--------|---------|----------|----------|-----------|------------|-----------|
|                  | h     | Acquisti | Vendite  | Prezzi | Delta   | Idro     | PV       | P.ggi     | Term.co    | Eolico    |
| Validi           | 8760  | 8760     | 8760     | 8760   | 8760    | 8760     | 8760     | 8760      | 8760       | 8760      |
| N Mancanti       | 0     | 0        | 0        | 0      | 0       | 0        | 0        | 0         | 0          | 0         |
| Media            |       | 2,2669   | 2,1880   | ,0931  | -,0789  | 4,2663   | 6,0946   | -4,9572   | 170,6353   | 22,2438   |
| E.S. della media |       | ,00416   | ,00400   | ,00044 | ,00125  | ,07097   | ,10283   | ,09481    | ,37095     | ,22156    |
| Mediana          |       | 2,2670   | 2,1840   | ,0900  | -,1000  | 1,1180   | ,0000    | ,0000     | 168,8610   | 15,6520   |
| Moda             |       | 2,53     | 2,43     | ,17    | -,10    | ,09      | ,00      | ,00       | 161,77     | ,95       |
| Deviazione std.  |       | ,38975   | ,37481   | ,04128 | ,11729  | 6,64217  | 9,62456  | 8,87355   | 34,71917   | 20,73649  |
| Varianza         |       | ,152     | ,140     | ,002   | ,014    | 44,118   | 92,632   | 78,740    | 1205,421   | 430,002   |
| Intervallo       | 23,00 | 1,89     | 1,86     | ,18    | ,55     | 41,37    | 40,51    | 41,11     | 201,76     | 100,10    |
| Minimo           | 1,00  | 1,38     | 1,30     | ,01    | -,30    | ,00      | ,00      | -41,11    | 73,27      | ,00       |
| Massimo          | 24,00 | 3,27     | 3,17     | ,19    | ,25     | 41,37    | 40,51    | ,00       | 275,03     | 100,10    |
| Somma            |       | 19857,74 | 19166,63 | 815,59 | -691,11 | 37372,59 | 53388,71 | -43424,80 | 1494765,44 | 194855,58 |

Tabella 12.3 - Statistiche descrittive DataSet anno 2011.

| Statistiche      |       |          |          |        |         |          |           |           |            |           |
|------------------|-------|----------|----------|--------|---------|----------|-----------|-----------|------------|-----------|
|                  | h     | Acquisti | Vendite  | Prezzi | Delta   | Idro     | PV        | P.ggi     | Term.co    | Eolico    |
| Validi           | 8784  | 8784     | 8784     | 8784   | 8784    | 8784     | 8784      | 8784      | 8784       | 8784      |
| N Mancanti       | 0     | 0        | 0        | 0      | 0       | 0        | 0         | 0         | 0          | 0         |
| Media            |       | 2,2783   | 2,1664   | ,0950  | -,1119  | 4,3380   | 12,7206   | -4,4022   | 156,9489   | 28,7470   |
| E.S. della media |       | ,00436   | ,00461   | ,00040 | ,00130  | ,07358   | ,19242    | ,08606    | ,37570     | ,28317    |
| Mediana          |       | 2,2840   | 2,1860   | ,0940  | -,1000  | 1,3760   | ,0860     | ,0000     | 154,0260   | 20,2960   |
| Moda             |       | 2,28     | 2,14     | ,17    | -,10    | ,09      | ,00       | ,00       | 151,27     | 1,98      |
| Deviazione std.  |       | ,40834   | ,43212   | ,03761 | ,12204  | 6,89644  | 18,03435  | 8,06596   | 35,21207   | 26,53998  |
| Varianza         |       | ,167     | ,187     | ,001   | ,015    | 47,561   | 325,238   | 65,060    | 1239,890   | 704,371   |
| Intervallo       | 23,00 | 2,05     | 2,24     | ,30    | ,55     | 42,74    | 60,63     | 42,48     | 260,92     | 199,35    |
| Minimo           | 1,00  | 1,39     | 1,12     | ,00    | -,30    | ,00      | ,00       | -42,48    | 57,88      | ,00       |
| Massimo          | 24,00 | 3,45     | 3,36     | ,30    | ,25     | 42,74    | 60,63     | ,00       | 318,80     | 199,35    |
| Somma            |       | 20012,65 | 19029,29 | 834,20 | -983,36 | 38104,97 | 111737,56 | -38668,70 | 1378639,56 | 252513,72 |

Tabella 12.4 - Statistiche descrittive DataSet anno 2012.

| Statistiche      |       |          |          |        |          |          |           |           |            |           |
|------------------|-------|----------|----------|--------|----------|----------|-----------|-----------|------------|-----------|
|                  | h     | Acquisti | Vendite  | Prezzi | Delta    | Idro     | PV        | P.ggi     | Term.co    | Eolico    |
| Validi           | 8760  | 8760     | 8760     | 8760   | 8760     | 8760     | 8760      | 8760      | 8760       | 8760      |
| N Mancanti       | 0     | 0        | 0        | 0      | 0        | 0        | 0         | 0         | 0          | 0         |
| Media            |       | 2,2068   | 2,0704   | ,0920  | -,1364   | 4,1521   | 15,7655   | -3,9787   | 150,9679   | 30,1200   |
| E.S. della media |       | ,00418   | ,00469   | ,00041 | ,00127   | ,07129   | ,23276    | ,08011    | ,36320     | ,29779    |
| Mediana          |       | 2,2060   | 2,1050   | ,0838  | -,1000   | 1,3760   | ,1720     | ,0000     | 147,0600   | 20,2960   |
| Moda             |       | 2,14     | 2,14     | ,06    | -,10     | ,09      | ,00       | ,00       | 138,55     | 9,03      |
| Deviazione std.  |       | ,39102   | ,43884   | ,03792 | ,11896   | 6,67204  | 21,78548  | 7,49774   | 33,99405   | 27,87205  |
| Varianza         |       | ,153     | ,193     | ,001   | ,014     | 44,516   | 474,607   | 56,216    | 1155,596   | 776,851   |
| Intervallo       | 23,00 | 2,05     | 2,18     | ,22    | ,55      | 43,00    | 72,84     | 40,42     | 220,42     | 129,86    |
| Minimo           | 1,00  | 1,28     | 1,04     | ,00    | -,30     | ,00      | ,00       | -40,42    | 80,07      | ,09       |
| Massimo          | 24,00 | 3,33     | 3,23     | ,22    | ,25      | 43,00    | 72,84     | ,00       | 300,48     | 129,95    |
| Somma            |       | 19331,79 | 18136,72 | 805,93 | -1195,07 | 36372,75 | 138105,51 | -34853,35 | 1322479,10 | 263850,79 |

Tabella 12.5 - Statistiche descrittive DataSet anno 2013.

| Statistiche      |           |          |          |        |          |          |           |           |            |           |
|------------------|-----------|----------|----------|--------|----------|----------|-----------|-----------|------------|-----------|
|                  | h         | Acquisti | Vendite  | Prezzi | Delta    | Idro     | PV        | P.ggi     | Term.co    | Eolico    |
| Validi           | 8760      | 8760     | 8760     | 8760   | 8760     | 8760     | 8760      | 8760      | 8760       | 8760      |
| N Mancanti       | 0         | 0        | 0        | 0      | 0        | 0        | 0         | 0         | 0          | 0         |
| Media            | 12,5000   | 2,0605   | 1,9259   | ,0809  | -,1346   | 4,0677   | 17,2506   | -4,1474   | 142,4273   | 30,1505   |
| E.S. della media | ,07396    | ,00378   | ,00440   | ,00038 | ,00130   | ,06948   | ,24970    | ,07885    | ,33366     | ,29938    |
| Mediana          | 12,5000   | 2,0790   | 1,9750   | ,0766  | -,1000   | 1,2040   | ,2580     | ,0000     | 138,9760   | 19,9520   |
| Moda             | 1,00      | 1,75     | 2,10     | ,06    | -,10     | ,09      | ,00       | ,00       | 134,25     | 1,29      |
| Deviazione std.  | 6,92258   | ,35416   | ,41141   | ,03600 | ,12133   | 6,50312  | 23,37080  | 7,38021   | 31,22894   | 28,02038  |
| Varianza         | 47,922    | ,125     | ,169     | ,001   | ,015     | 42,291   | 546,194   | 54,468    | 975,247    | 785,142   |
| Intervallo       | 23,00     | 1,90     | 2,15     | ,18    | ,55      | 47,04    | 79,38     | 43,17     | 213,11     | 134,68    |
| Minimo           | 1,00      | 1,14     | ,84      | ,00    | -,30     | ,00      | ,00       | -43,17    | 57,96      | ,00       |
| Massimo          | 24,00     | 3,03     | 2,99     | ,18    | ,25      | 47,04    | 79,38     | ,00       | 271,07     | 134,68    |
| Somma            | 109500,00 | 18049,69 | 16870,74 | 708,88 | -1178,95 | 35632,64 | 151114,90 | -36331,13 | 1247662,76 | 264118,00 |

Tabella 12.6 - Statistiche descrittive DataSet anno 2014.

È quindi possibile comparare i grafici delle frequenze per i singoli fattori in studio.

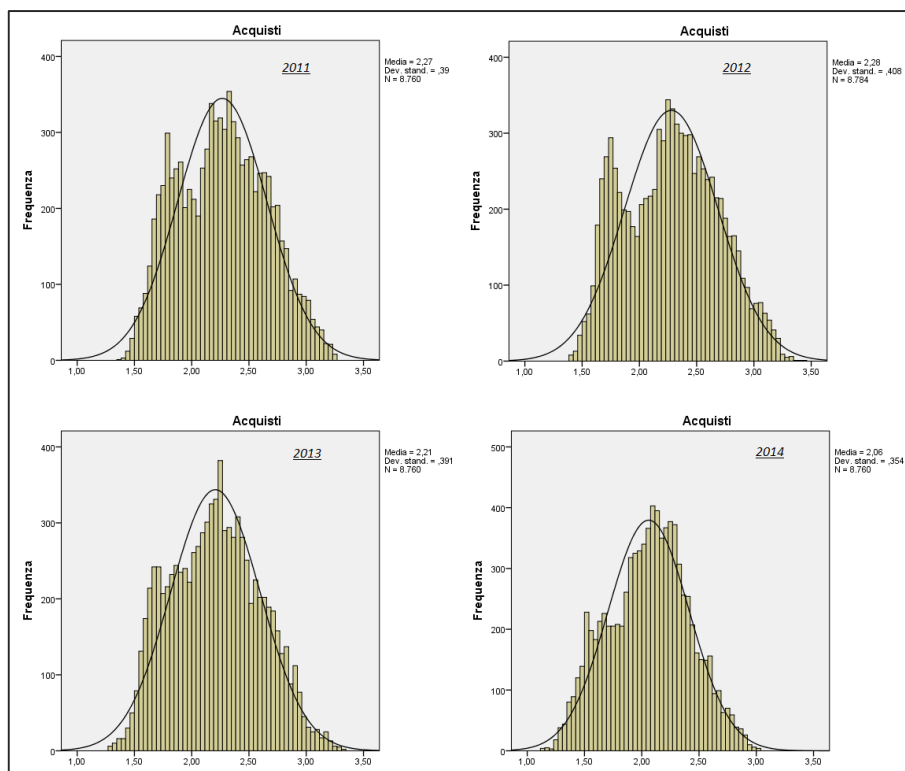


Figura 12.11 - Frequenza dei ktep di acquisto nel MGP per gli anni 2011,2012, 2013 e 2014.

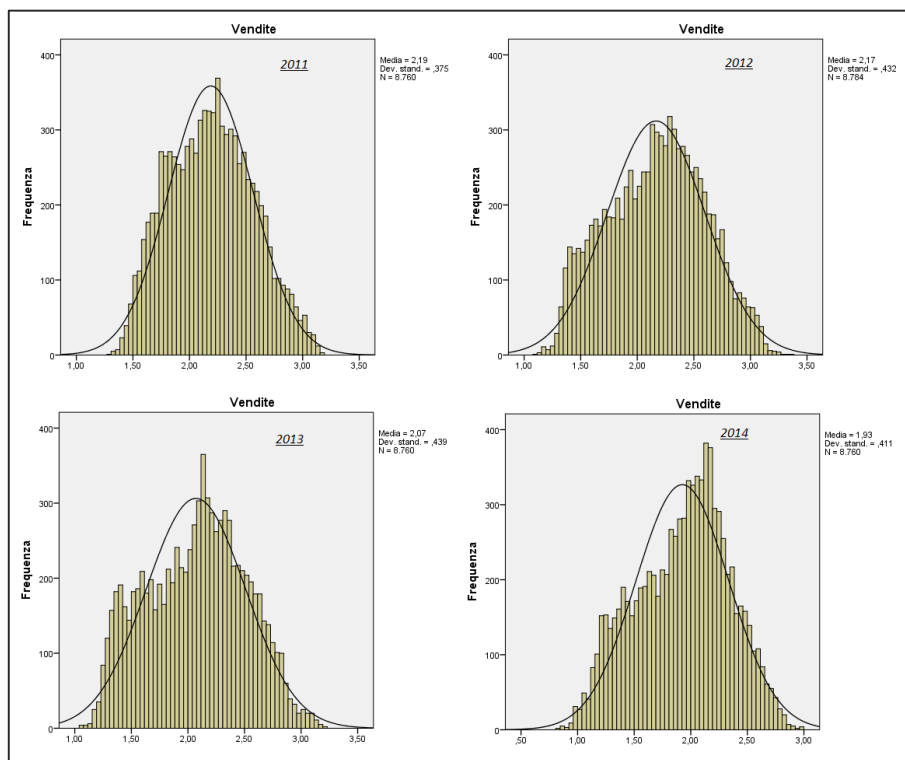


Figura 12.12 - Frequenza dei ktep di vendite nel MGP per gli anni 2011,2012, 2013 e 2014.

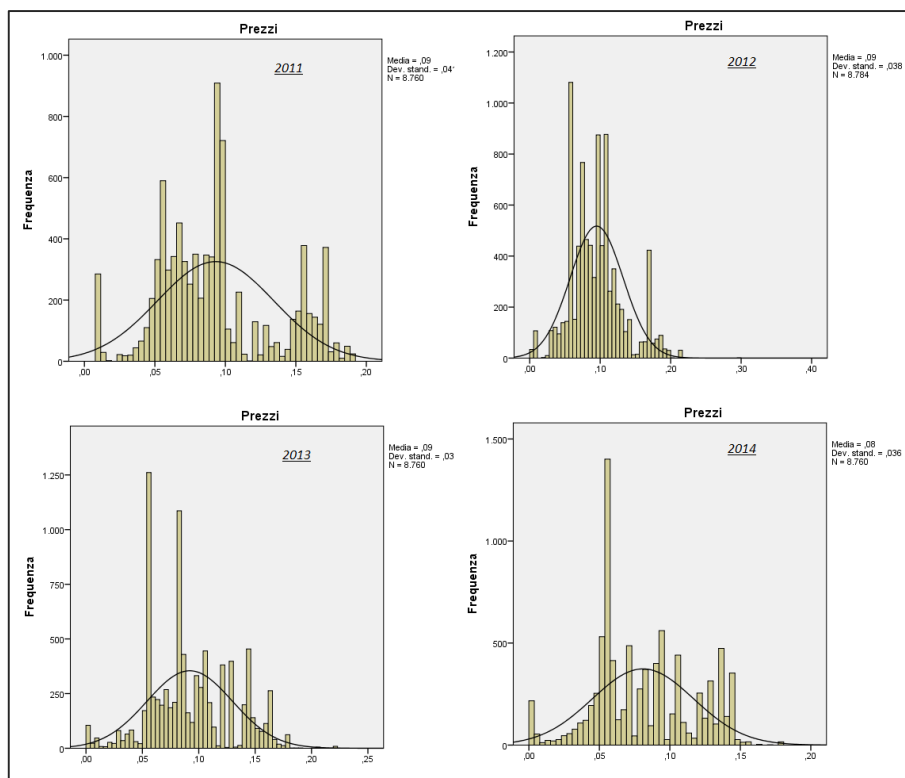


Figura 12.13 - Frequenza dei prezzi in €/tep d nel MGP per gli anni 2011,2012, 2013 e 2014.

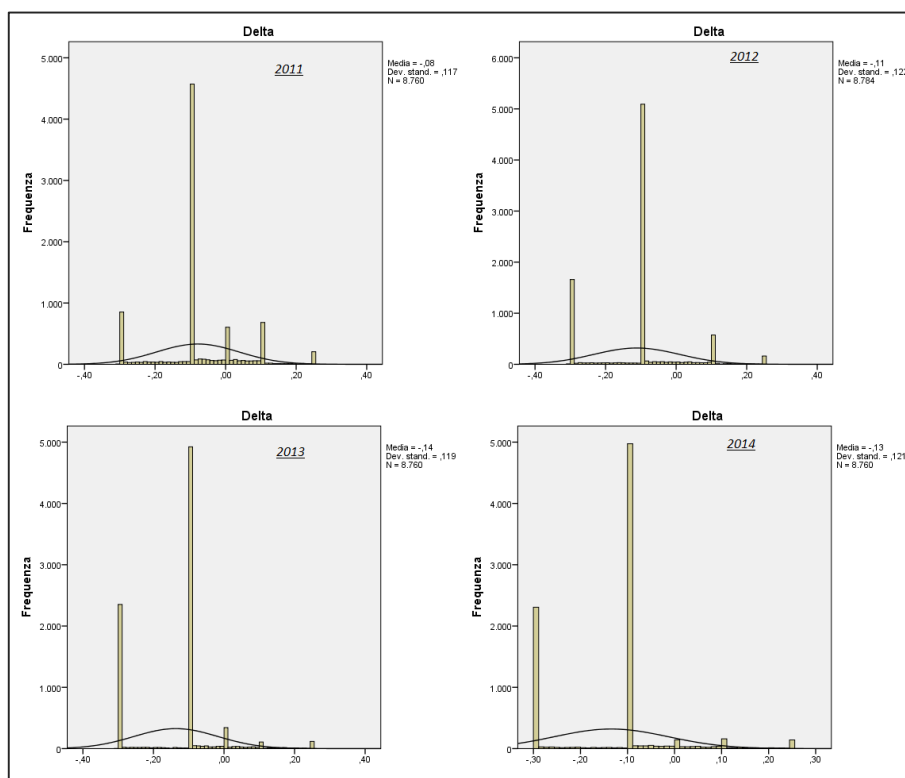


Figura 12.14 - Frequenza del parametro delta in tep nel MGP per gli anni 2011,2012, 2013 e 2014.



Di seguito sono riportati i plot dei valori di produzione in tep orari degli impianti. Attraverso lo studio dei grafici è possibile vedere eventuali stagionalità e valori anomali.

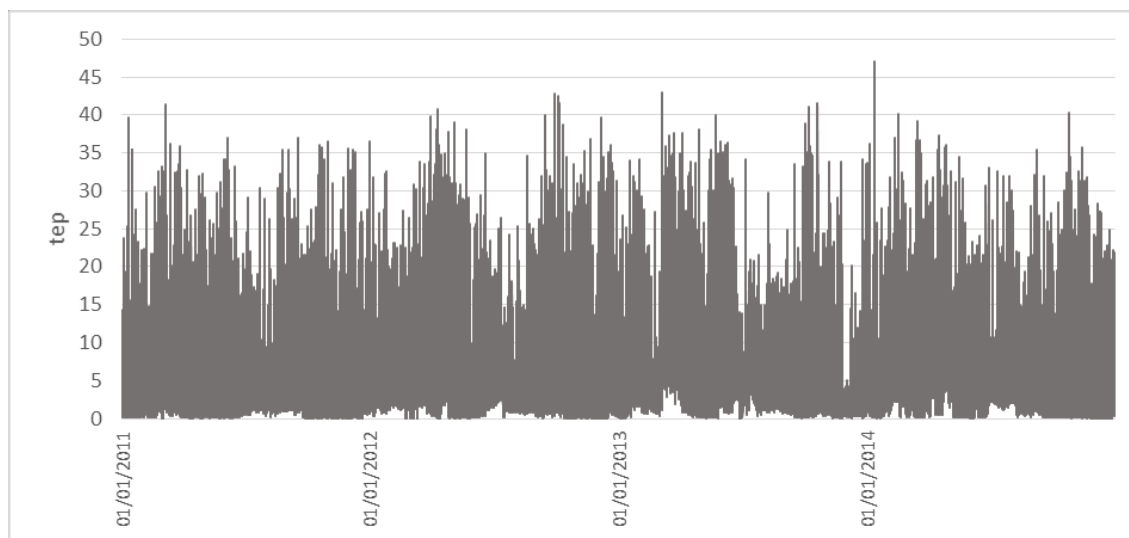


Figura 12.15 - Plot produzione idroelettrica dal 2011 al 2014.

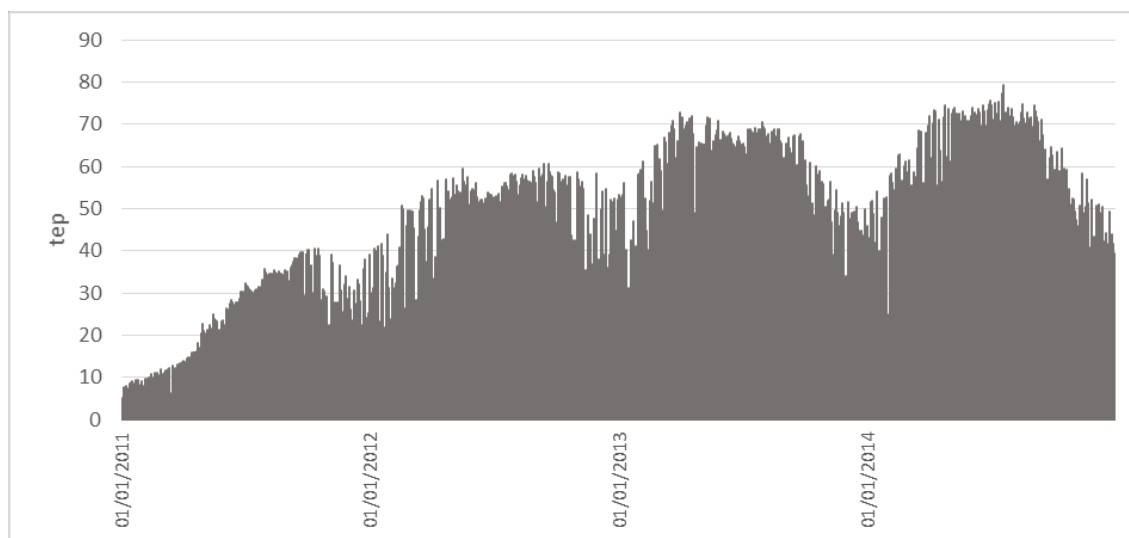


Figura 12.16 - Plot produzione fotovoltaica dal 2011 al 2014.

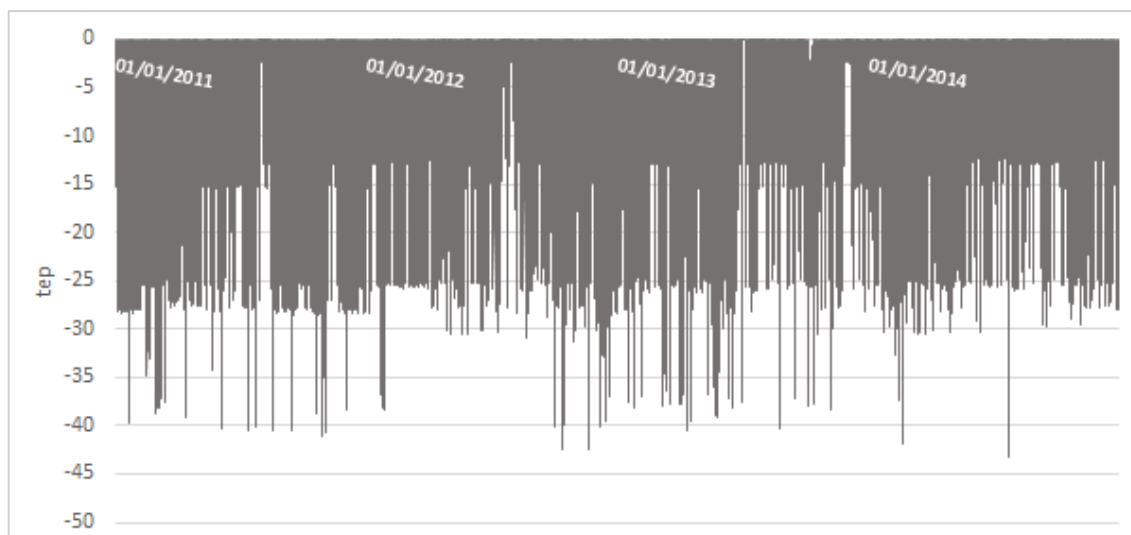


Figura 12.17 - Plot energia richiesta dagli impianti a pompaggio dal 2011 al 2014.

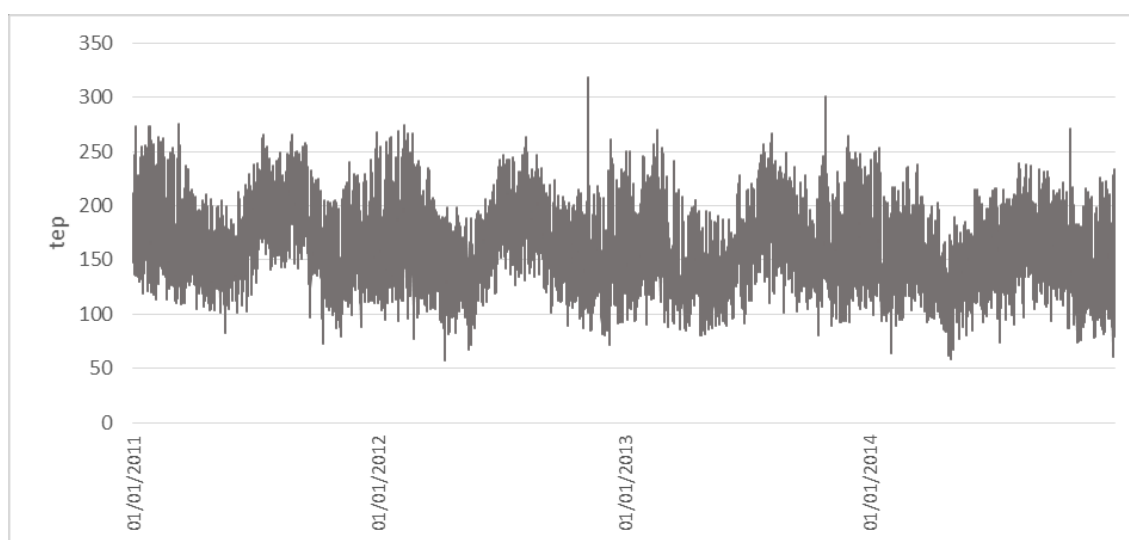


Figura 12.18 - Plot produzione termoelettrica dal 2011 al 2014.

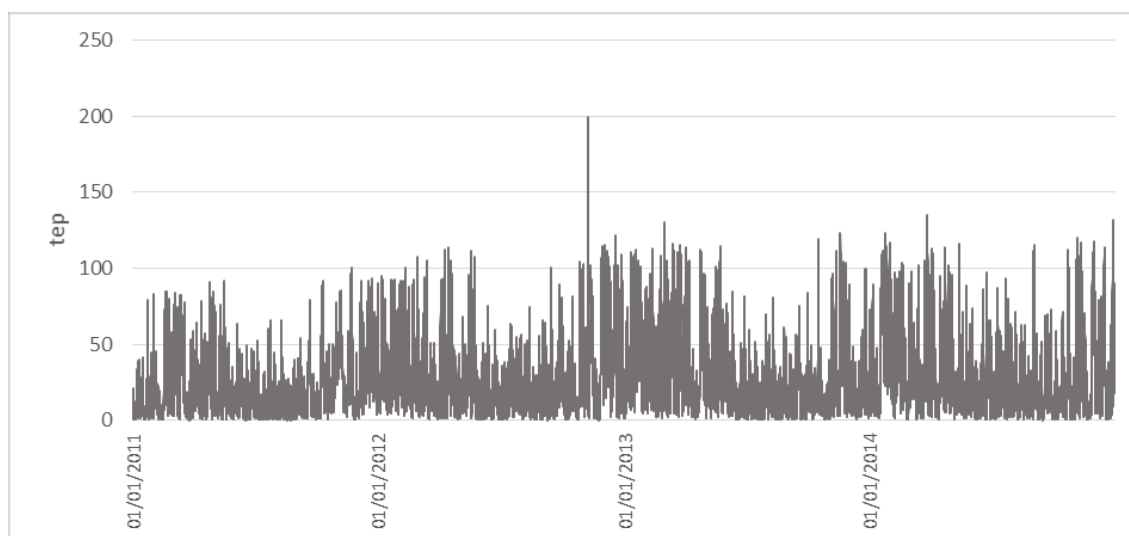


Figura 12.19 - Plot produzione eolica dal 2011 al 2014.

Attraverso i dati riportati nella Figura 12.20 è mostrato un quadro riassuntivo con le frequenze del parametro “Delta” in relazione all’aumento delle ore in cui le offerte si assestano sul MGP con prezzo nullo.

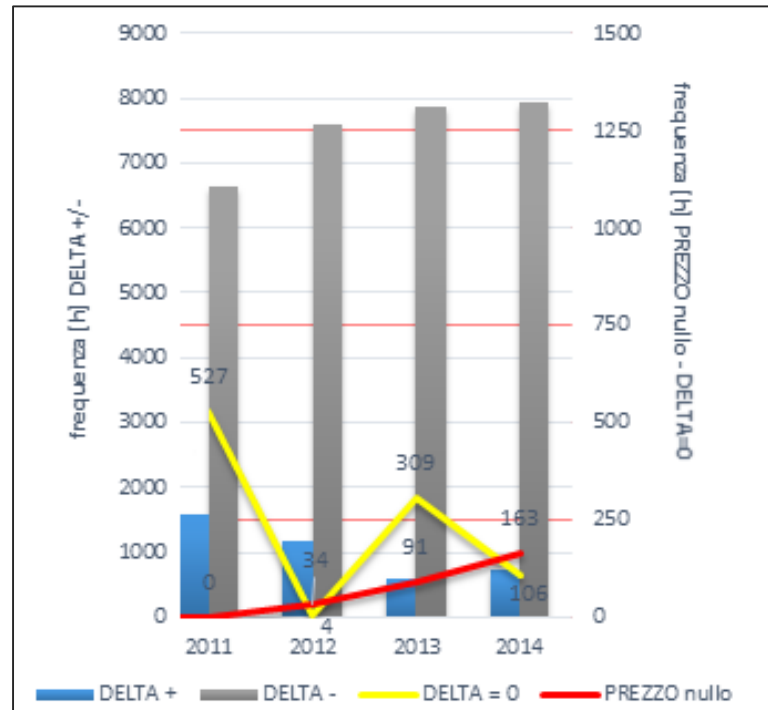


Figura 12.20 - Frequenza Delta-Prezzi nulli.

Si evidenzia in particolare come il numero di ore con offerte a Prezzo nullo (linea in rosso) siano aumentate notevolmente. Ciò risponde, essenzialmente, all’influenza delle FER, che immettendosi sul MGP con costo marginale pari a zero, consentono di ridurre il prezzo zonale dell’energia elettrica, con ricadute anche sul PUN, oltre che spostare anche le ore in cui ciò accade.

Quest’ultimo fenomeno prende il nome di peak shaving e consiste nell’aumento spropositato del prezzo dell’energia soprattutto nelle prime ore serali, quando per via del profilo di fabbisogno, e per la scarsità di apporto delle FER-E in queste ore (soprattutto fotovoltaico), le offerte in immissione vengono coperte prevalentemente dal termoelettrico a prezzi notevolmente più alti per recuperare i profitti di cui si è privato durante le ore diurne a causa del criterio di merit order che le ha escluse fino alle ore serali.

Interessante è la correlazione riguardante le ore in cui si verificano le offerte a prezzo nullo in funzione dei vari anni (Figura 12.21).

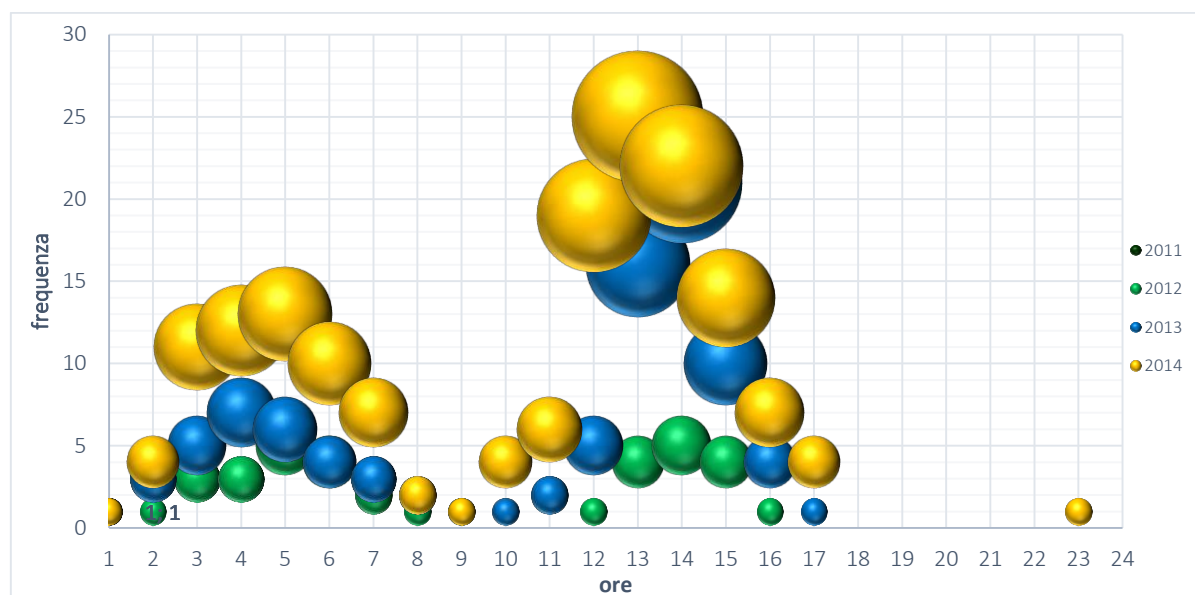


Figura 12.21 - Frequenza ore a prezzo nullo.

In particolare è da notare che la frequenza delle ore con prezzo di offerta nullo degli anni 2013 e 2014 si concentra nella fascia oraria 12-15. Questo si può spiegare chiamando in causa la natura stesse delle FER, le quali (vedi fotovoltaico ed eolico) concentrano maggiormente la loro produttività nel corso delle ore centrali della giornata.

Ciò naturalmente è in controtendenza con quanto accadeva prima dell'esplosione delle FER; infatti la minor richiesta notturna ha portato per un certo tempo a tariffe orarie mediamente più basse. Negli ultimi anni si assiste all'inversione di questo fenomeno proprio perché il tradizionale termoelettrico cerca di recuperare i margini di profitto che ha perso durante le ore diurne in cui è stato escluso per via della presenza delle FER-E.

## 12.4 Modello di Previsione Neural Network “Prezzo zonale”

Il modello sviluppato con il software GMDH Shell basato sulle Neural Network GMDH ha come scopo la definizione del prezzo zonale ponendo come input della rete neurale i dati relativi allo sbilanciamento delle vendite e delle offerte (parametro Delta), l'energia generata dalle FER-E. e la componente oraria. Nella Figura 12.22 si riporto lo schema della rete utilizzata.

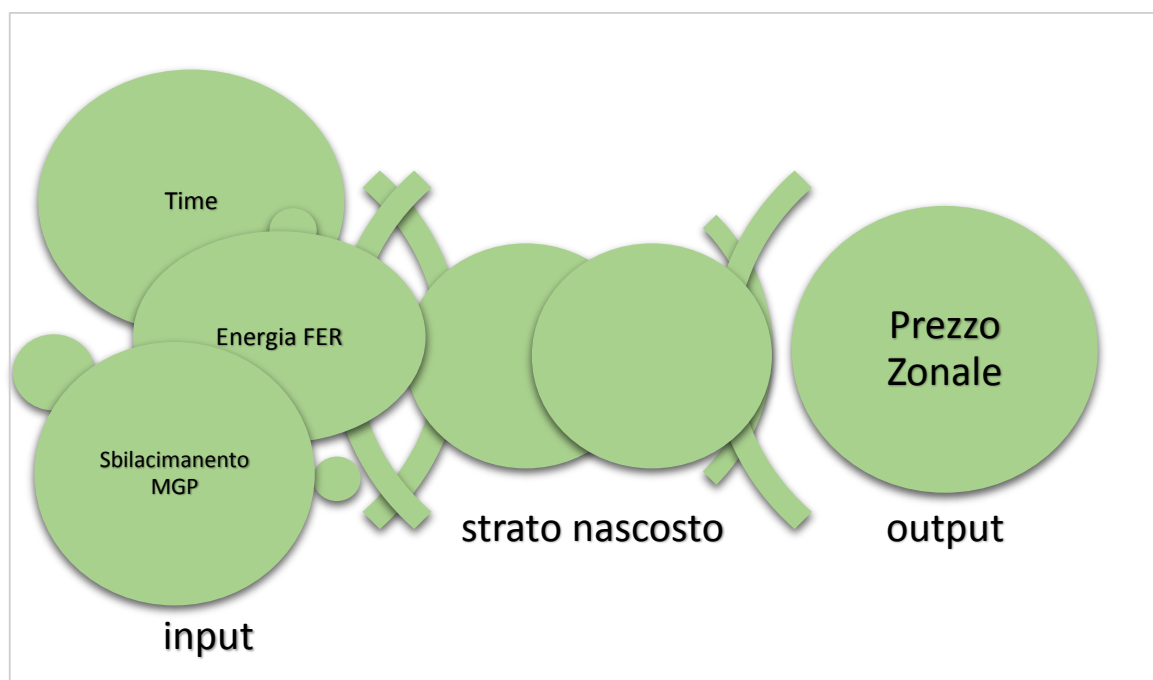


Figura 12.22 – Diagramma neurale - Previsione "Prezzo Zonale".

Attraverso l'iterazione di più settaggi si è scelto quello che coniugasse basso errore e velocità computazionale. Sono state quindi definite le seguenti impostazioni nel pannello Solver (Tabella 12.7).

|                       |                               |
|-----------------------|-------------------------------|
| Reorder observations  | Odd/even                      |
| Validation strategy   | k-fold validation             |
| Number of folds       | 2                             |
| Validation criterion  | RMSE                          |
| Variables ranking     | No                            |
| Core algorithm        | GMDH neural network           |
| Neuron function       | $a + x_i + x_i * x_j + x_i^2$ |
| Max. number of layers | 5                             |
| Initial layer width   | 5                             |
| Time series mode      | ✓                             |

Tabella 12.7 – Settings solver – Modello di previsione "Prezzo Zonale".

In particolare evidenziamo il criterio di validazione RMSE, che rappresenta la deviazione standard del campione delle differenze tra i valori previsti ed i valori osservati, il core algorithm (GMDH neural network) che rappresenta l'algoritmo di apprendimento statistico, la funzione neurale di attivazione del singolo neurone ( $a + x_i + x_i * x_j + x_i^2$ ) con la quale si ottiene il plot del model fit e della predictions, il massimo numero di strati (5) come limite superiore per il numero di strati di rete creati dall'algoritmo,

e la larghezza dello strato iniziale (5) che definisce quanti neuroni vengono aggiunti alla serie di ingresso ad ogni nuovo livello.

Avviando il processo di training e di validation si sono quindi ottenuti i seguenti risultati (Tabella 12.8).

| Postprocessed results                     | Model fit | Predictions |
|---|-----------|-------------|
| Number of observations                    | 28051     | 7013        |
| Max. negative error                       | -0,1548   | -0,1153     |
| Max. positive error                       | 0,1191    | 0,0956      |
| Normalized Mean absolute error (nMAE)     | 6,67 %    | 6,69 %      |
| Normalized Root mean square error (nRMSE) | 8,83 %    | 8,82 %      |
| Residual sum                              | 0,0312    | -1,0384     |
| Standard deviation of residuals           | 0,0265    | 0,0264      |
| Coefficient of determination ( $R^2$ )    | 0,5297    | 0,5293      |
| Correlation                               | 0,7278    | 0,7275      |

Tabella 12.8 – Accuracy – Modello di previsione “Offerta di Vendita” con Neural Network.

Il calcolo del MAE e del RMSE in forma normalizzata permette di superare l’errore che si genera con il MAPE ed il RMSPE per via della prossimità al valore 0 dei dati di target.

La figura 12.23 rappresenta il plot del model fit e della predizione.

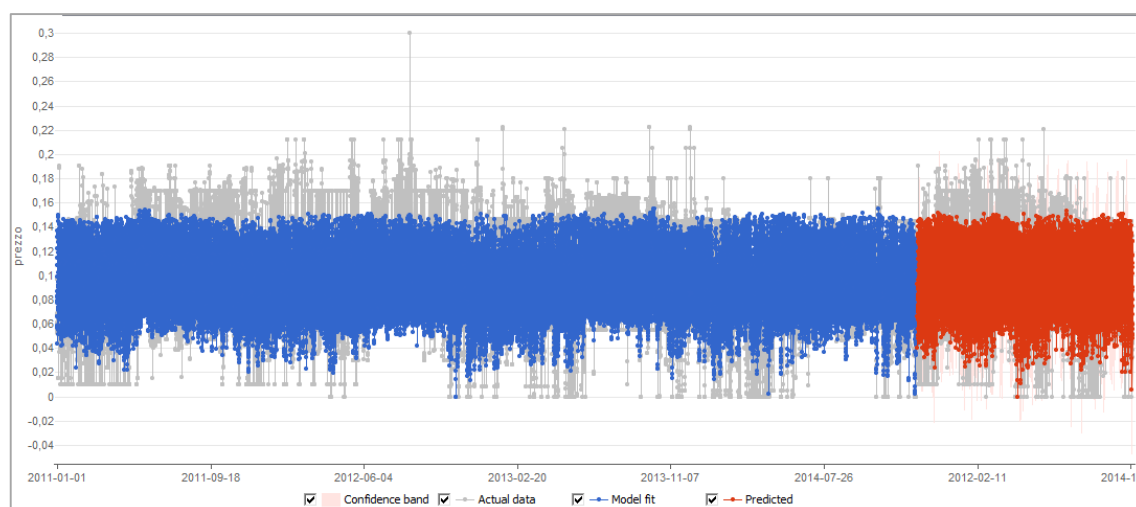


Figura 12.23 – Plot model fit e previsione “Prezzo Zonale” con Neural Network.

Nelle figure 12.24, 12.25 e 12.26 sono mostrati rispettivamente il “residuals” del model fit e della predictions, l’autocorrelazione e la frequenza del valore residuo.

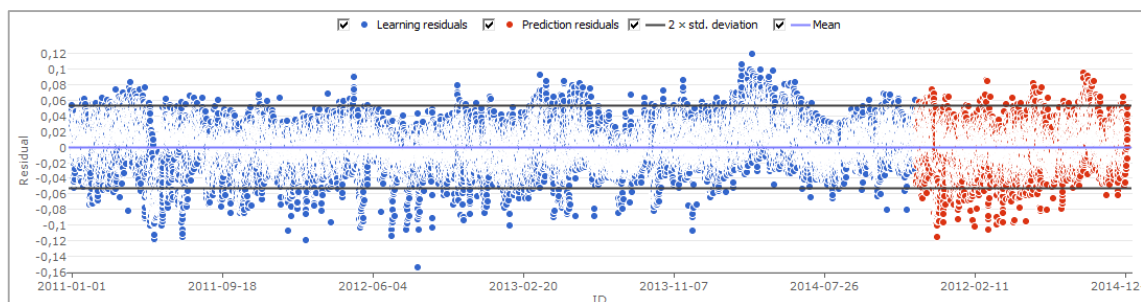


Figura 12.24 – Plot residui previsione “Prezzo Zonale” con Neural Network.

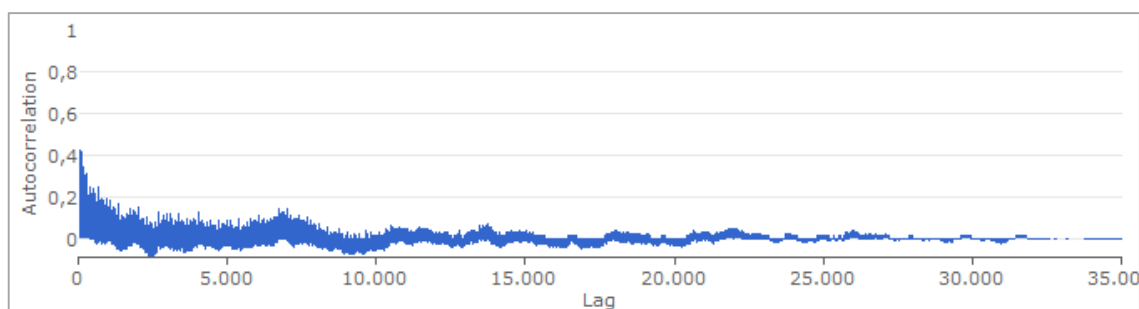


Figura 12.25 – Autocorrelazione previsione “Prezzo Zonale” con Neural Network.

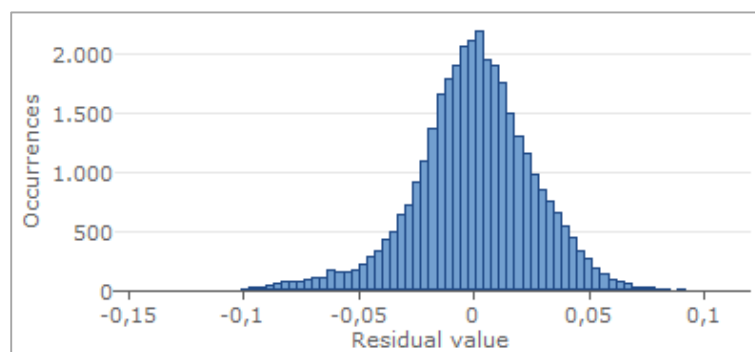


Figura 12.24 – Frequenza residui previsione “Prezzo Zonale” con Neural Network.

Di seguito si riporta la funzione generata per la predizione.

### **Prezzo zonale**

$$Y1 = 0.00573193 + N2*0.569975 + N2*N3*131.016 - N2^2*64.3909 + N3*0.308993 - N3^2*65.8374$$

$$N3 = -0.0153916 + N4*0.744474 + N4*N6*4.4882 - N4^2*2.96444 + N6*0.526119 - N6^2*2.47799$$

$$N6 = -0.0540006 + N8*0.373239 + N8*N10*0.528091 + N8^2*2.28734 + N10*1.35058 - N10^2*4.20567$$

$$N10 = 0.0891666 + Idro*0.00378612 - Idro*Wind*3.01523e-06 - Idro^2*6.44836e-05 - Eolico*0.000366671 - Eolico^2*2.24809e-07$$

$$N8 = 0.0987065 - h*0.00348779 + h*\text{"delta, cubert"}*0.000152785 + h^2*0.000228619 - \text{"delta, cubert"}*0.0362637 - \text{"delta, cubert"}^2*0.0955097$$

$$N2 = -0.00686877 + N9*0.538031 + N9*N4*3.29031 - N9^2*2.3775 + N4*0.363976 + N4^2*0.938066$$

$$N4 = -0.0881166 - Eolico*0.000369057 + Eolico*N7*0.000977359 - Eolico^2*1.24643e-06 + N7*3.1185 - N7^2*10.7066$$

$$N7 = 0.0796226 - Ora*0.00455359 - Ora*PV*0.000211488 + Ora^2*0.000308269 + PV*0.00359932 - Photovoltaic^2*1.07181e-05$$

$$N9 = 0.0934993 - \text{"delta, cubert"}*0.0420768 + \text{"delta, cubert"}*Idro*0.000293609 - \text{"delta, cubert"}^2*0.111802 + Idro*0.0031885 - Idro^2*5.4838e-05$$

## **12.5 Modello di Previsione serie storiche “Prezzo zonale”**

Il modello sviluppato con il software GMDH Shell basato su *Stepwise mixed selection* di GMDH ha come scopo la definizione del prezzo zonale ponendo come input i dati storici orari del prezzo zonale dall'anno 2013 all'anno 2014. Il SetData è stato ridotto rispetto al modello regressivo ai soli anni 2013 e 2014 per via della complessità computazionale e del tempo di iterazione.

Dopo diverse iterazioni con più settaggi si è scelto quello che coniugasse basso errore e velocità computazionale. Sono state quindi definite le seguenti impostazioni nel pannello Solver (Tabella 12.7).

|                           |                          |
|---------------------------|--------------------------|
| Reorder observations      | Pseud-random             |
| Validation strategy       | k-fold validation        |
| Number of folds           | 2                        |
| Validation criterion      | RMSE balance             |
| Variables ranking         | By correlation           |
| Drop variables after rank | 25                       |
| Core algorithm            | Stepwise mixed selection |
| Neuron function           | $x_i * x_j$              |
| Limit model complexitt to | 200                      |
| Return set of best models | 100                      |
| Time series mode          | ✓                        |

Tabella 12.9 – Settings solver – Modello di previsione “Prezzo Zonale”.



Avviando il processo di training con la previsione dei successivi 8760 valori, si sono quindi ottenuti i seguenti risultati (Tabella 12.10).

| Postprocessed results                     | Model fit              |
|---|------------------------|
| Number of observations                    | 8760                   |
| Max. negative error                       | -0,0731                |
| Max. positive error                       | 0,0831                 |
| Normalized Mean absolute error (nMAE)     | 6,67 %                 |
| Normalized Root mean square error (nRMSE) | 8,83 %                 |
| Residual sum                              | $7,50 \text{ e}^{-12}$ |
| Standard deviation of residuals           | 0,0202                 |
| Coefficient of determination ( $R^2$ )    | 0,6861                 |
| Correlation                               | 0,8634                 |

Tabella 12.10 – Accuracy – Modello di previsione “Prezzo Zonale”.

La figura 12.25 rappresenta il plot del model fit e della previsione.

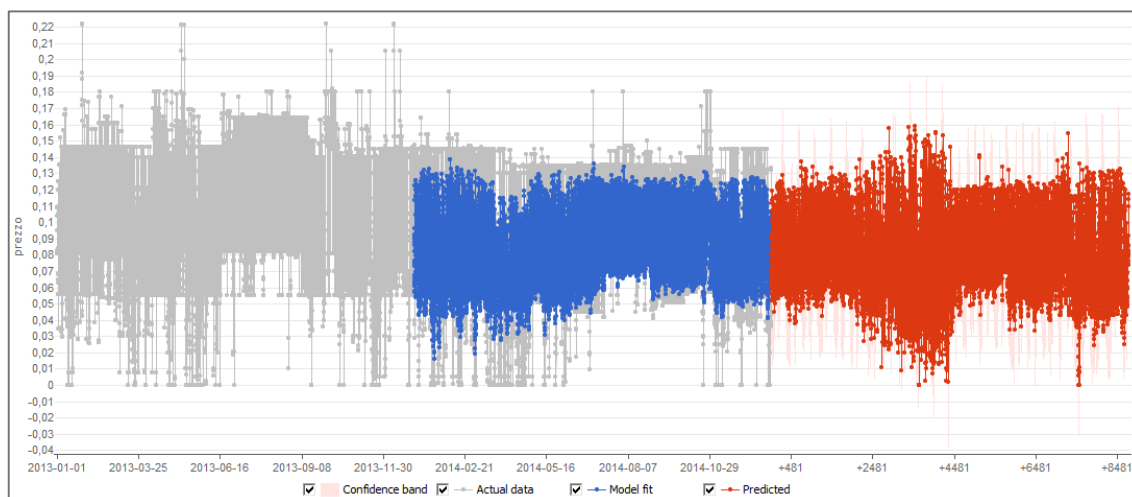


Figura 12.25 – Plot model fit e previsione “Prezzo Zonale”

## 12.6 Conclusioni

Dal confronto dei due modelli implementati si nota come il modello serie storiche dia, seppur di poco, migliori indici rispetto a quello basato su modello regressivo NN. Infatti, nella Tabella 10.11 si riscontrano valori di massimi errori negativi e positivi più bassi nel modello serie storiche, come la somma dei residui. Gli indici di correlazione ed il coefficiente  $R^2$  risultano invece più elevati a parità di nMAE ed nRMSE.

| Postprocessed results                     | NN      | Serie storiche        |
|---|---------|-----------------------|
| Number of observations                    | 7013    | 8760                  |
| Max. negative error                       | -0,1153 | -0,0731               |
| Max. positive error                       | 0,0956  | 0,0831                |
| Normalized Mean absolute error (nMAE)     | 6,69 %  | 6,67 %                |
| Normalized Root mean square error (nRMSE) | 8,82 %  | 8,83 %                |
| Residual sum                              | -1,0384 | 7,50 e <sup>-12</sup> |
| Standard deviation of residuals           | 0,0264  | 0,0202                |
| Coefficient of determination ( $R^2$ )    | 0,5293  | 0,6861                |
| Correlation                               | 0,7275  | 0,8634                |

Tabella 12.11 – Confronto modelli di previsione “Prezzo Zonale”.

La quasi uguaglianza dei due modelli in termini di indicatori di bontà, è essenzialmente dovuta dalla buona organizzazione del DataSet, (serie continue e pluriennali) e dall’elevata inerzia di un sistema energetico su scala regionale che generalmente presenta caratteristiche di stagionalità e ripetitività. Tali caratteristiche rendono la predizione con modelli di serie storiche molto accurate e con bassi errori, facendo sì che tali modelli siano competitivi rispetto a quelli NN basati su regressioni.

## CAPITOLO 13 *Caso studio*

### *edificio uso ufficio*

L'Unione europea (U.E.) fin dalla sua costituzione è sempre stata attenta alle tematiche ambientali [48] e dell'approvvigionamento energetico [49], infatti negli anni ha identificato varie azioni da attuare [50, 51] per raggiungere standard specifici ed stringenti per la riduzione delle emissioni di CO<sub>2</sub> [52], il risparmio energetico e la produzione di energia da fonte rinnovabile [53]. Nel 2013, è stato fatto un ulteriore passo in avanti con la definizione delle politiche per il clima e l'energia con target 2030 [54], che delinea la continuazione dell'importante percorso per il risparmio energetico e l'efficienza energetica che gli Stati membri hanno già iniziato. La nuova strategia propone come obiettivi che ogni Stato membro dovrà raggiungere:

- la riduzione delle emissioni del 40% rispetto ai livelli del 1990;
- la produzione di almeno il 27% di energie rinnovabili nell'U.E..

Nel contesto attuale dell'Unione, gli interventi necessari per raggiungere gli obiettivi di politica energetica colpiscono, soprattutto, i principali settori economici dei singoli Stati membri [55], in particolare l'industria, il settore civile e dei trasporti. Nel settore civile, tuttavia, risulta strategico valutare il contributo delle attività che si svolgono all'interno di edifici uso ufficio e che hanno una ricaduta diretta nei consumi energetici. Infatti, la regolarità della domanda di energia, sia giornaliera ma anche mensili e stagionali, rende le valutazioni energetiche prevedibili. Pertanto, politiche energetiche locali che perseguono obiettivi di efficienza energetica per questo tipo di edifici possono portare a risultati concreti nel breve e medio termine.

Con lo studio condotto sono quindi stati analizzati i fattori che influenzano il rendimento energetico degli edifici uso ufficio e le caratteristiche energetiche dell'involucro. Dai risultati ottenuti possono essere evinti importanti considerazioni per una migliore adattamento delle politiche energetiche, degli standard di progettazione e delle strategie operative.

I fattori più importanti sono i parametri climatici, perché rappresentano le condizioni al contorno per la progettazione ed influenzano il comportamento transitorio dell'involucro edilizio nel corso della sua vita di servizio [56]. Tra i parametri climatici, l'indicatore Gradi-Giorno (DD) può essere utilizzato per prevedere la richiesta di energia per il riscaldamento e raffrescamento degli ambienti. Molti studi utilizzano DD per le analisi energetiche locali e per la predizione della domanda di energia [57]. Ad

esempio, in Lee, K. et altri [58], gli autori hanno studiato la domanda energetica dell'edificio legato ai cambiamenti climatici, valutando richieste di elettricità e combustibile per il riscaldamento nel Commonwealth e nel Massachusetts. I risultati della ricerca hanno quindi mostrato notevoli cambiamenti nel consumo energetico dell'edificio e del proprio mix energetico. Kim e Suh in [59] hanno usato la procedura di ASHRAE, basandola sui DD di riscaldamento (HDD) e DD di raffrescamento (CDD), per stimare la domanda di energia in un edificio sulla base di scenari che prevedono i cambiamenti climatici nei prossimi anni. In Yi-Ling, H. et altri [60], gli autori hanno usato le relazioni tra il consumo di energia e HDD e CDD per indagare il consumo per riscaldamento nei mesi invernali e consumo per raffrescamento nei mesi estivi.

Per quanto espresso, è certamente importante indagare su questa stretta relazione tra indicatori climatici e carichi interni. Nel presente caso studio si sono quindi sviluppate correlazioni semplici, basate su metodi regressivi, per ottenere il valore dei carichi termici per il riscaldamento al variare dell'indice HDD. Tali correlazioni sono state sviluppate per sette diversi paesi europei, attraverso dati empirici e simulazioni con il software TRNSYS [61] di un edificio tipo uso ufficio caratterizzato dai parametri costruttivi relativi agli standard energetici di ogni paese. Questo ha permesso lo sviluppo di sette correlazioni validate per i sette paesi e tre correlazioni valide per le zone continentali (settentrionale, centrale e sud Europa), con valori elevati del parametro  $R^2$ .

### **13.1 La bibliografia sugli edifici uso ufficio**

Gli edifici adibiti ad uffici possono essere suddivisi in tre categorie:

- ufficio personale (unico occupante);
- ufficio del gruppo di lavoro (da due a sei posti di lavoro);
- ufficio openspace (sette o più posti di lavoro).

Alcuni rapporti, come il Rapporto sui risultati del progetto - EL-TERZIARY [62], hanno documentato come il consumo di energia per gli uffici è legato essenzialmente agli usi finali di elettricità e calore, e che non vi è grande differenza tra gli Stati per quanto riguarda l'uso finale di energia elettrica. Infatti, il consumo di energia elettrica è tipicamente legato alle apparecchiature per ufficio e di illuminazione che hanno un consumo specifico in kWh/m<sup>2</sup> piuttosto simile in tutti i paesi industrializzati. Alcune differenze sono però rinvenibili nei paesi con significativi valori di CDD, a causa del quale si riscontrano consumi per il condizionamento, ventilazione e refrigerazione, pari a circa il 18% della domanda totale di energia elettrica. Naturalmente, la domanda di energia per il riscaldamento varia secondo le condizioni atmosferiche, ma è anche influenzata dalle caratteristiche costruttive degli edifici e in particolare dal coefficiente globale di scambio termico (valore U) delle pareti, finestre,

pavimenti e tetti. Tuttavia, va sottolineato che, oltre del valore U, il tasso di occupazione e la destinazione d'uso di un edificio hanno un effetto significativo sul rendimento energetico di un singolo edificio [63].

### 13.2 Gli stati oggetto dello studio

Lo studio condotto, al fine di ottimizzare il numero di simulazioni da effettuare e nel contempo fornire risultati rappresentativi del contesto meteorologico rappresentativo europeo, si è concentrato su solo sette Stati, ognuno caratteristico da un ben preciso clima e per i quali sono stati riscontrati in bibliografia dati empirici confrontabili.

Schematicamente, l'Europa è caratterizzata da cinque diversi ambienti climatici (Figura 13.1):

- clima subartico o alta montagna;
- clima continentale (che è diviso in clima continentale freddo e clima continentale freddo);
- Atlantic clima;
- Mountain clima;
- clima mediterraneo.

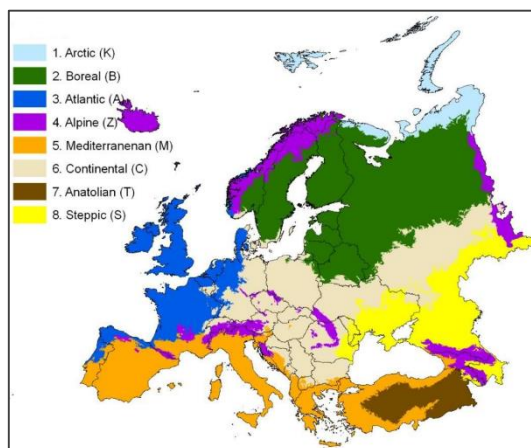


Figura 13.1 - Classificazione dei climi europei.

Come precedentemente descritto, il valore DD è un indicatore che può essere utilizzato per la valutazione e l'analisi del tempo correlato al consumo di energia di un edificio. Per limitare lo studio a un gruppo selezionato di Stati, che sono rappresentativi di differenti condizioni geo-climatiche e delle caratteristiche costruttive nell'UE, una classificazione iniziale è stato realizzato utilizzando di HDD e CDD indicatori, forniti da Eurostat [64] e riportati nella Tabella 13.1.

| European Countries |                | HDD  | CDD  | Solar irradiance<br>[kW/m <sup>2</sup> year] |
|--------------------|----------------|------|------|--|
| 1                  | Finland        | 5078 | 27   | 869  |
| 2                  | Sweden         | 4735 | 33   | 876  |
| 3                  | Estonia        | 4396 | 39   | 964  |
| 4                  | Latvia         | 4212 | 38   | 998  |
| 5                  | Lithuania      | 4116 | 37   | 998  |
| 6                  | Luxembourg     | 3534 | 63   | 1037   |
| 7                  | Poland         | 3494 | 83   | 1011   |
| 8                  | Denmark        | 3482 | 32   | 976  |
| 9                  | C. Republic    | 3268 | 125  | 1022   |
| 10                 | Slovakia       | 3100 | 201  | 1088   |
| 11                 | Germany        | 3095 | 91   | 1000   |
| 12                 | Romania        | 3088 | 290  | 1351   |
| 13                 | Austria        | 3060 | 158  | 1077   |
| 14                 | Slovenia       | 2904 | 207  | 1205   |
| 15                 | Hungary        | 2834 | 254  | 1226   |
| 16                 | Ireland        | 2806 | 0    | 952  |
| 17                 | Bulgaria       | 2706 | 304  | 1460   |
| 18                 | United Kingdom | 2669 | 42   | 949  |
| 19                 | Netherlands    | 2669 | 70   | 988  |
| 20                 | Belgium        | 2619 | 110  | 923  |
| 21                 | France         | 2265 | 202  | 1278   |
| 22                 | Italy          | 1442 | 561  | 1424   |
| 23                 | Spain          | 1161 | 748  | 1588   |
| 24                 | Greece         | 1155 | 958  | 1460   |
| 25                 | Portugal       | 886  | 325  | 1570   |
| 26                 | Malta          | 634  | 916  | 1763   |
| 27                 | Cyprus         | 512  | 1290 | 1738   |

Tabella 13.1 - HDD, CDD e radiazione solare degli Stati europei.

Tracciando i valori della Tabella 13.1 in Figura 13.2, è possibile osservare che i dati dell'indicatore HDD tendono ad essere inversamente proporzionale al valore di irraggiamento solare, mentre i dati dell'indicatore CDD sono abbastanza proporzionale.

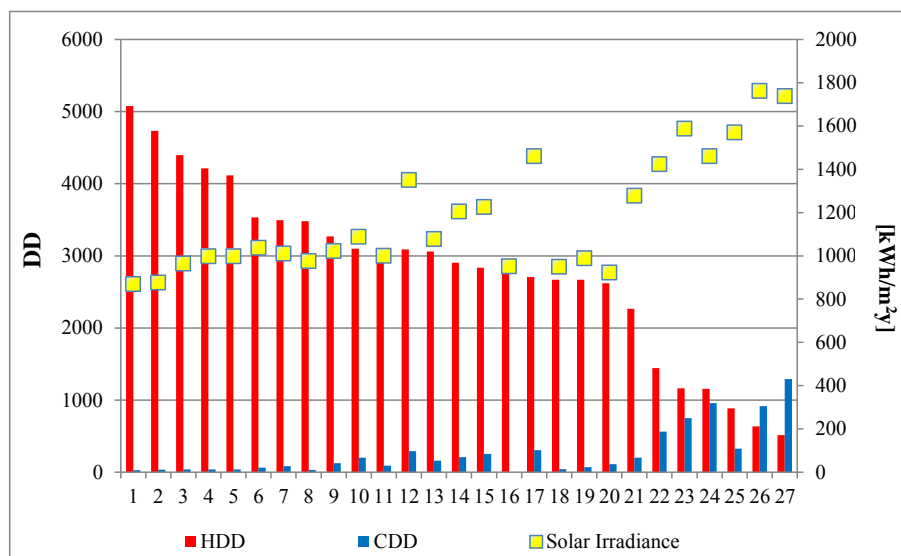


Figura 13.2 - HDD, CDD e radiazione solare degli Stati europei.

Per semplificare l'analisi e, allo stesso tempo, per preservare il significato dello studio, sono stati scelti solo pochi paesi tra il gruppo UE. La scelta ha quindi teso ad analizzare almeno un paese

rappresentativo delle diverse aree climatiche del territorio europeo. La Figura 13.3 identifica le nazioni che sono state selezionate in verde: Svezia, Germania, Regno Unito, Belgio, Francia, Italia e Spagna.

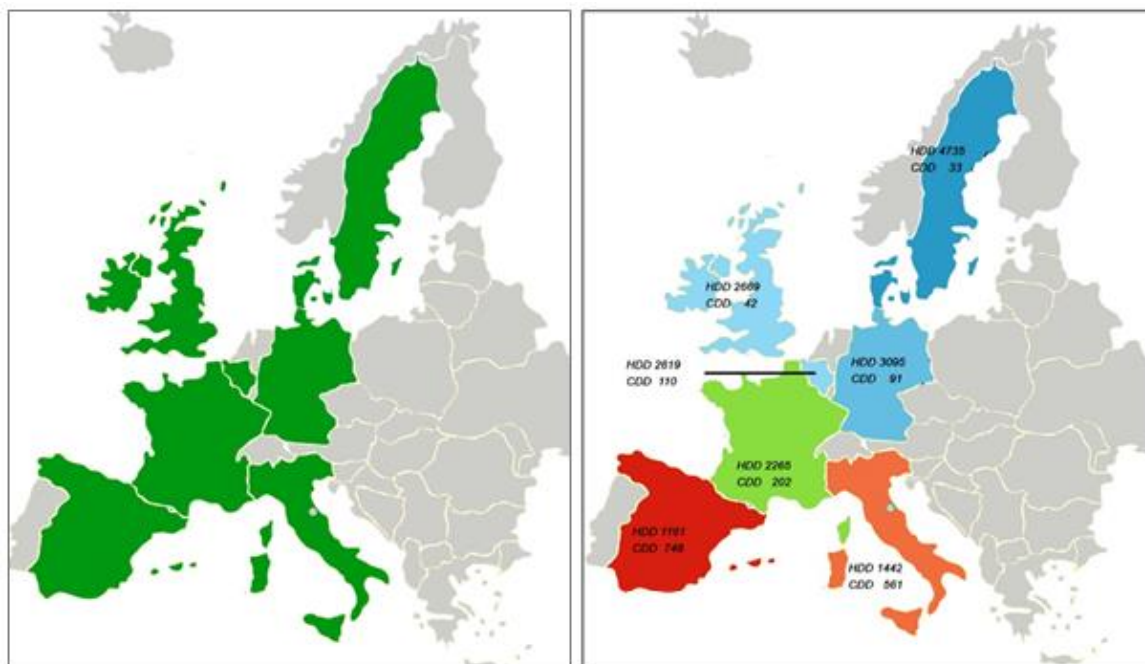


Figura 13.3 - Paesi oggetto dello studio e relativi valori di HDD e CDD.

| European Countries |                | HDD  | CDD | Solar irradiance<br>[kW/m <sup>2</sup> year] |
|--------------------|----------------|------|-----|--|
| 1                  | Sweden         | 4735 | 33  | 876  |
| 2                  | Germany        | 3095 | 91  | 1000   |
| 3                  | United Kingdom | 2669 | 42  | 949  |
| 4                  | Belgium        | 2619 | 110 | 923  |
| 5                  | France         | 2265 | 202 | 1278   |
| 6                  | Italy          | 1442 | 561 | 1424   |
| 7                  | Spain          | 1161 | 748 | 1588   |

Tabella 13.2 - Valori di HDD, CDD e radiazione solare per gli Stati oggetto dello studio.

Grazie ai dati forniti da Eurostat è stato possibile determinare un intervallo di valori di HDD massimo e minimo [64] (Figura 13.4).

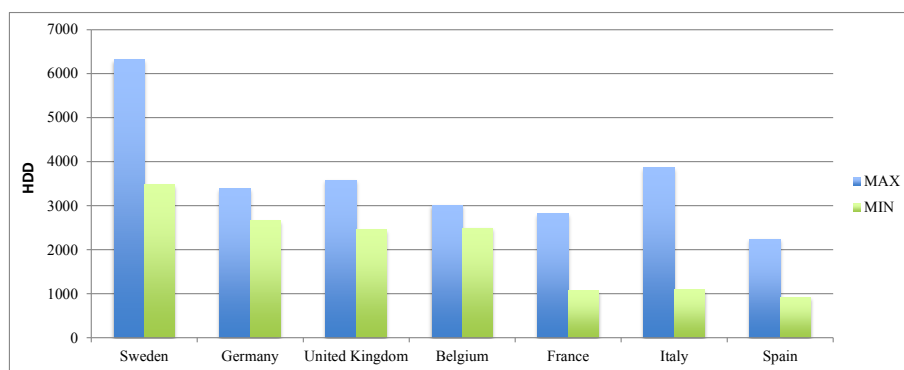


Figura 13.4 - HDD massimo e minimo degli stati oggetto dello studio.

L'analisi dei dati di sintesi iniziale consente l'identificazione di cinque diversi range di HDD; il numero di punti di dati in ogni range di HDD, è mostrato in Tabella 13.3.

|             | HDD>4000 | 3000<HDD<3999 | 2000<HDD<2999 | 1000<HDD<1999 | HDD<1000 |
|-------------|----------|---------------|---------------|---------------|----------|
| Svezia      | 5        | 3             | 0             | 0             | 0        |
| Germania    | 0        | 25            | 16            | 0             | 0        |
| Regno Unito | 0        | 32            | 33            | 0             | 0        |
| Belgio      | 0        | 1             | 10            | 0             | 0        |
| Francia     | 0        | 0             | 20            | 3             | 0        |
| Italia      | 0        | 3             | 5             | 13            | 0        |
| Spagna      | 0        | 0             | 2             | 13            | 3        |

Tabella - 13.3 HDD range.

L'analisi dei dati tracciati in Tabella 13.3 mostra che:

- più del 60% dei punti dati della Svezia sono caratterizzati da valori superiori a HDD 4000;
- il 60% dei punti di dati di Germania e il 50% del Regno Unito sono tra i valori 4000 e 3000 di HDD;
- circa il 90% dei punti dati in Belgio e Francia sono tra il 3000 e il 2000 HDD;
- il 60% dei punti dati in Italia e circa il 70% della Spagna sono tra il 2000 e il 1000 HDD.

In ultimo, attraverso la bibliografia riportata in Tabella 13.4 è stato possibile ricavare i dati empirici di energia per riscaldamento richiesta dagli edifici ad uso ufficio nei singoli Stati.



| European Countries | Period    | kWh/m <sup>2</sup> | Reference |
|--------------------|-----------|--------------------|-----------|
| Sweden             | average   | 58,3               | [55]      |
| Germany            | average   | 74                 | [56-60]   |
| United Kingdom     | ante 1990 | 175                | [61]      |
|                    | post 1990 | 95                 |           |
| Belgium            | ante 1975 | 140                | [62]      |
|                    | post 1975 | 95                 |           |
| France             | average   | 115                | [63]      |
| Italy              | North     | 92                 | [64,65]   |
|                    | Centre    | 23,5               |           |
|                    | South     | 6,5                |           |
| Spain              | post 1990 | 15                 | [66]      |

Tabella 13.4 - Dati empirici sull'energia termica richiesta dagli edifici uso ufficio negli Stati oggetto dello studio.

A sua volta per ogni Stato sono state scelte tre città che potessero caratterizzare il range di HDD tipico dello Stato, Tabella 13.5.

| City               | HDD    |
|--------------------|--------|
| BE-Malbedy         | 3214,4 |
| BE-Ostende         | 2602,5 |
| BE-St,Hubert       | 3225,4 |
| DE-Fichtelberg     | 5968,1 |
| DE-Frankfurt       | 2930,7 |
| DE-Hof             | 3491,3 |
| ES-Madrid          | 2159,7 |
| ES-Salamanca       | 2626,5 |
| ES-Sevilla         | 810,8  |
| FR-Bordeaux        | 2118,5 |
| FR-Bourges         | 2972,6 |
| FR-Nice            | 1444,2 |
| IT-Porto Empedocle | 691,2  |
| IT-Sestriere       | 6618,0 |
| IT-Venezia         | 2316,0 |
| SE-Kiruna          | 7730,6 |
| SE-Lund            | 3299,6 |
| SE-Umea            | 5301,0 |
| UK-Aviemore        | 3510,3 |
| UK-Birmingham      | 2820,1 |
| UK-Camborne        | 2224,3 |

Tabella 13.5 - Città prese ad esame nello studio.

### 13.3 Il modello di simulazione

Individuati gli Stati oggetto dello studio, si è proceduto a costruire il modello di simulare su TRNSYS. Il modello è quindi stato caratterizzato dai parametri di guadagni interni, ventilazione e portata d'aria dettati dalla norma EN ISO 13790: 2008.

In particolare, per la determinazione dei valori relativi ai carichi interni gratuiti si è utilizzato l'allegato G della norma, che specifica i valori di carico interno di default per gli edifici per uffici (Tabella 13.6 e Tabella 13.7). Infine, la portata d'aria di ventilazione è stata valutata seguente allegato C della EN ISO 13790.

| Giorno                                  | Ore           | Stanze ufficio<br>[W/m <sup>2</sup> ] | Altre stanze<br>[W/m <sup>2</sup> ] |
|---|---------------|---------------------------------------|-------------------------------------|
|   |               |                                       |                                     |
| Lunedì-Venerdì                          | 07.00 – 17.00 | 20,0                                  | 8,0                                 |
|   | 17.00 – 23.00 | 2,0                                   | 1,0                                 |
|   | 23.00 – 07.00 | 2,0                                   | 1,0                                 |
|   | Average       | 9,50                                  | 3,92                                |
| Sabato-Domenica                         | 07.00 – 17.00 | 2,0                                   | 1,0                                 |
|   | 17.00 – 23.00 | 2,0                                   | 1,0                                 |
|   | 23.00 – 07.00 | 2,0                                   | 1,0                                 |
|   | Average       | 2,0                                   | 1,0                                 |
| Valore medio                            |               | 7,4                                   | 3,1                                 |
| Valore medio delle sole apparecchiature |               | 3,0                                   | -                                   |

Tabella 13.6 - Somma dei contributi gratuiti di occupanti e apparecchiature [65].

| Giorno          | Ore           | Persone<br>[W/m <sup>2</sup> ] | Apparecchiature<br>[W/m <sup>2</sup> ] |
|-----------------|---------------|--------------------------------|--|
|                 |               |                                |  |
| Lunedì-Venerdì  | 07.00 – 17.00 | 9,0                            | 6,2                                    |
|                 | 17.00 – 23.00 | 0                              | 1,6                                    |
|                 | 23.00 – 07.00 | 0                              | 1,6                                    |
|                 | Average       | 3,7                            | 3,5                                    |
| Sabato-Domenica | 07.00 – 17.00 | 0                              | 1,0                                    |
|                 | 17.00 – 23.00 | 0                              | 1,0                                    |
|                 | 23.00 – 07.00 | 0                              | 1,0                                    |
|                 | Average       | 0                              | 1,0                                    |
| Valore medio    |               | 2,6                            | 2,7                                    |

Tabella 13.7 - Contributi gratuiti di occupanti e apparecchiature.

A sua volta l'involucro è stato caratterizzato tenendo conto dei dati riportati dalla Insulation Manufacturers Association (EURIMA) [66], ed organizzando i dati reperiti in report relativi a

caratteristiche strutturali per ogni paese. Di conseguenza, i valori U, relativi a superfici opache, come pareti, pavimenti e tetti, sono stati classificati per HDD in Tabella 13.8.

|             |                  | HDD>4000 | 3000<HDD<3999 | 2000<HDD<2999 | 1000<HDD<1999 | HDD<1000 |
|-------------|------------------|----------|---------------|---------------|---------------|----------|
| Svezia      | <i>parete</i>    | 0,18     | 0,18          | -             | -             | -        |
|             | <i>pavimento</i> | 0,15     | 0,15          |               |               |          |
|             | <i>tetto</i>     | <0,13    | <0,13         |               |               |          |
| Germania    | <i>parete</i>    | -        | 0,25          | 0,25          | -             | -        |
|             | <i>pavimento</i> |          | 0,40          | 0,40          |               |          |
|             | <i>tetto</i>     |          | 0,20          | 0,20          |               |          |
| Regno Unito | <i>parete</i>    | -        | 0,25          | 0,25          | -             | -        |
|             | <i>pavimento</i> |          | 0,20          | 0,20          |               |          |
|             | <i>tetto</i>     |          | 0,18          | 0,18          |               |          |
| Belgio      | <i>parete</i>    | -        | > 0,6         | > 0,6         | -             | -        |
|             | <i>pavimento</i> |          | > 0,6         | > 0,6         |               |          |
|             | <i>tetto</i>     |          | 0,40          | 0,40          |               |          |
| Francia     | <i>parete</i>    | -        | -             | 0,35          | 0,35          | -        |
|             | <i>pavimento</i> |          |               | 0,25          | 0,35          |          |
|             | <i>tetto</i>     |          |               | 0,20          | 0,25          |          |
| Italia      | <i>parete</i>    | -        | 0,4           | 0,4           | > 0,6         | -        |
|             | <i>pavimento</i> |          | 0,45          | 0,45          | > 0,6         |          |
|             | <i>tetto</i>     |          | 0,40          | 0,40          | >0,6          |          |
| Spagna      | <i>parete</i>    | -        | -             | > 0,6         | > 0,6         | > 0,6    |
|             | <i>pavimento</i> |          |               | > 0,6         | > 0,6         | > 0,6    |
|             | <i>tetto</i>     |          |               | 0,40          | 0,40          | 0,40     |

Tabella 13.8 - Valori medi di trasmittanza per pareti pavimenti e tetti per valore di HDD nei singoli stati.

I valori della trasmittanza delle superfici vetrate utilizzate sono riportate in Tabella 13.9

| Countries   | U-value delle finestre<br>tipo<br>[W/m <sup>2</sup> K] |      |
|-------------|--|------|
|             | from   | to   |
| Svezia      | 1,25   |      |
| Germania    | 1,25   |      |
| Regno Unito | 1,75   | 2,25 |
| Belgio      | 1,75   |      |
| Francia     | 1,75   |      |
| Italia      | 2,75   | 3,25 |
| Spagna      | 2,75   | 3,25 |

Tabella 13.9 - Valori U-Value delle finestre.

Per quanto riguarda i dati meteo, essi sono stati generati per ogni località oggetto dello studio (Tabella 13.5) utilizzando il software Meteonorm [67]. Il formato utilizzato presenta una gamma di fenomeni meteorologici in collaborazione con medie annuali che sono coerenti con le medie di lungo periodo per la località in questione.

Il periodo di riscaldamento è stato determinato per ogni città (Tabella 13.10). Sulla base della zona climatica è stata identificata una temperatura di soglia ( $T_{th}$ ). Il periodo di riscaldamento inizia dopo 3 giorni consecutivi che la differenza tra la temperatura di base e la temperatura esterna ( $T_{out}$ ) è inferiore al  $T_{th}$ , vale a dire:

$$T=(T_b-T_{out})\leq T_{th} \quad [95]$$

analogamente, il periodo di riscaldamento termina dopo 3 giorni consecutivi che la differenza di temperatura è:

$$T=(T_b-T_{out})>T_{th} \quad [96]$$

| Countries   | Città           | HDD     | Valori  | Periodo di riscaldamento |       |           |
|-------------|-----------------|---------|---------|--------------------------|-------|-----------|
|             |                 |         |         | da                       | a     | n. giorni |
| Svezia      | Kiruna          | 7730,60 | Massimo | Nessun limite            |       | 262       |
|             | Umea            | 5300,95 | Medio   | 01/09                    | 15/05 | 185       |
|             | Lund            | 3299,60 | Minimo  | 15/10                    | 15/05 | 153       |
| Germania    | Fichtelberg     | 5968,10 | Massimo | Nessun limite            |       | 262       |
|             | Hof             | 3491,25 | Medio   | 15/10                    | 15/05 | 153       |
|             | Francoforte     | 2930,71 | Minimo  | 15/10                    | 15/05 | 153       |
| Regno Unito | Aviemore        | 3510,30 | Massimo | 15/10                    | 15/05 | 153       |
|             | Birmingham      | 2820,14 | Medio   | 15/10                    | 15/05 | 153       |
|             | Camborne        | 2224,25 | Minimo  | 15/10                    | 15/05 | 153       |
| Belgio      | Malbedy         | 3214,35 | Massimo | 15/10                    | 15/05 | 153       |
|             | St. Hubert      | 3225,40 | Medio   | 15/10                    | 15/05 | 153       |
|             | Ostende         | 2602,47 | Minimo  | 15/10                    | 15/05 | 153       |
| Francia     | Belfort         | 2972,60 | Massimo | 15/10                    | 15/05 | 153       |
|             | Bordeaux        | 2118,51 | Medio   | 15/10                    | 15/05 | 153       |
|             | Nizza           | 1444,20 | Minimo  | 01/11                    | 30/04 | 153       |
| Italia      | Sestriere       | 6617,95 | Massimo | Nessun limite            |       | 262       |
|             | Venezia         | 2316,00 | Medio   | 15/10                    | 15/04 | 131       |
|             | Porto Empedocle | 686,85  | Minimo  | 01/12                    | 31/03 | 87        |
| Spagna      | Salamanca       | 2626,49 | Massimo | 15/10                    | 15/05 | 153       |
|             | Madrid          | 2159,65 | Medio   | 15/10                    | 15/05 | 153       |
|             | Seville         | 810,75  | Minimo  | 15/11                    | 15/03 | 85        |

Tabella 13.10 - Valori di HDD massimi, minimi e medi e periodi di riscaldamento delle città oggetto dello studio.

Per quanto riguarda le caratteristiche di forma, l'edificio scelto per questo studio è caratterizzato da una geometria regolare (Figura 13.4), con un tetto piano e pianta a C.

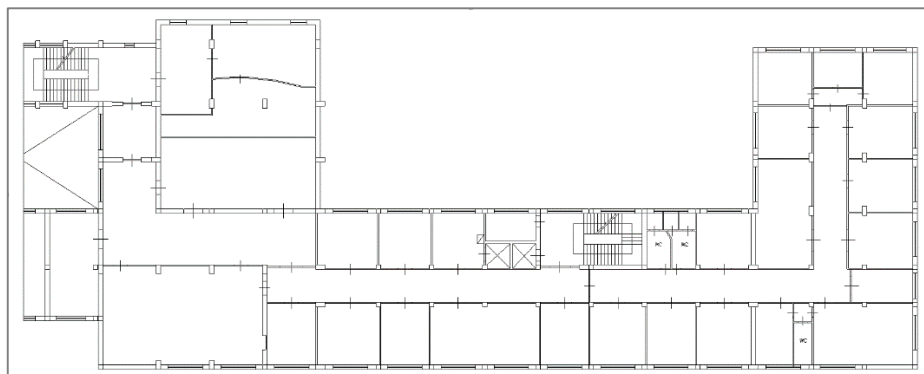


Figura 13.4 - Planimetria tipica di edificio tipo ufficio.

L'edificio si compone di cinque elevazioni; il piano seminterrato è concepito come un locale tecnico, mentre il piano terra è destinato all'uso archivio. Il primo, secondo e terzo piano contiene uffici, segretarie, sale biblioteca, amministrazione e del personale; ogni piano ha una superficie netta che varia da circa 1.100-1.500 m<sup>2</sup>. Figura 13.5 rappresentano i diversi punti di vista della costruzione.



Figura 13.5 - Rendering dei prospetti dell'edificio tipo ufficio.

Sulla base della destinazione d'uso, la posizione geometrica, esposizione, carichi elettrici e termici, il modello è stato suddiviso in 12 differenti zone termiche, Figura 13.6. Le zone termiche sono state definite come:

- Zona 1: Si estende per tre piani e si riferisce ai locali ascensore, scale e servizi igienici;
- Zona 2: è il piano sotterraneo ed è destinato a magazzino con basso tasso di occupazione;
- Zona 3: è l'intero terzo piano del palazzo;
- Zona 4, 5 e 12: uffici del primo e del secondo piano;
- Zona 6: sale conferenze al primo e al secondo piano;
- Zona 7: grande sala conferenze sala che si estende su due piani;
- Zona 8: corridoi che si estendono su tre piani;
- Zona 9: ingresso del palazzo, la reception e un piccolo ufficio;
- Zona 10, 11: uffici del piano terra.

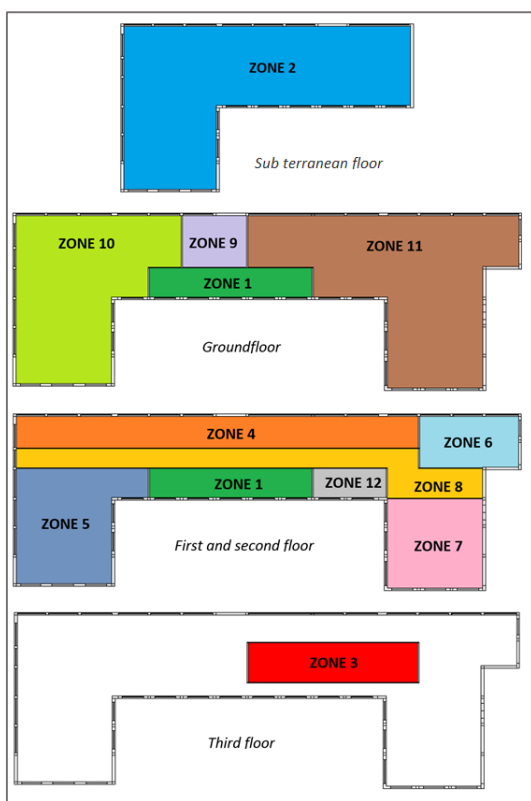


Figura 13.6 - Rappresentazione delle zone termiche dell'edificio tipo ufficio.

Per valutare il fabbisogno energetico, è stato quindi realizzato un modello dinamico su TRNSYS 17 [61] (Figura 13.7) che avesse specifiche condizioni meteorologiche esterne e tipiche caratteristiche strutturali.

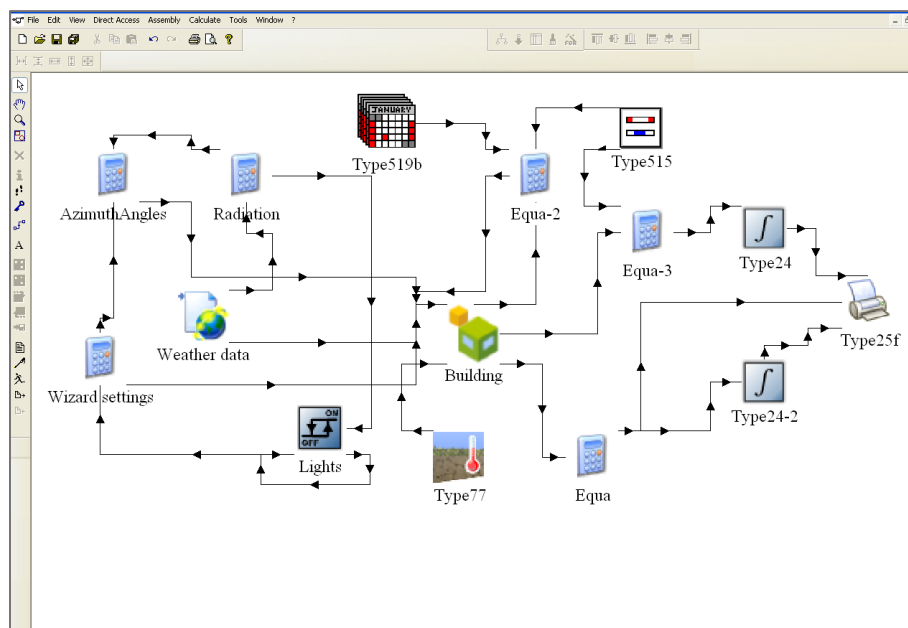


Figura 13.7 - Modello costruito su TRNSYS.

Infine, le caratteristiche termofisiche utilizzate sono riportate nella Tabella 13.11.

|             | HDD     | U-Value Finestre<br>[W/m²K] | m²finestre/m²pareti | m²tetto/m²pareti | U <sub>0</sub><br>[W/m²K] |
|-------------|---------|-----------------------------|---------------------|------------------|---------------------------|
| Svezia      | 4655,11 | 1,25                        | 0,48                | 0,63             | 0,476                     |
| Germania    | 3479,83 | 1,25                        | 0,64                | 0,90             | 0,656                     |
| Regno Unito | 2926,79 | 1,75                        | 0,64                | 0,90             | 0,805                     |
| Belgio      | 3075,38 | 1,75                        | 0,64                | 0,90             | 1,004                     |
| Francia     | 2767,02 | 1,75                        | 0,64                | 0,90             | 0,855                     |
| Italia      | 1608,36 | 2,75                        | 0,61                | 0,83             | 1,317                     |
| Spagna      | 2040,44 | 2,75                        | 0,61                | 0,83             | 1,390                     |

Tabella 13.11 - Valori termofisici utilizzati nello sviluppo del modello su TRNSYS.

L'ultima colonna della Tabella 13.11 presenta il valore U<sub>0</sub> generale, che è definita come l'indice termico unico che collega il fabbisogno energetico per il riscaldamento alle caratteristiche architettoniche dell'edificio. Il U<sub>0</sub> è data dall'equazione:

$$U_0 = \frac{\sum_i^n U_i S_i}{\sum_i^n S_i} \quad [97]$$

dove:

- $i$ : è la superficie esterna  $i$ -esimo, opaco e trasparente;
- $U_i$ : è valore  $U$  della superficie  $i$ -esima misurato in  $W / K / m^2$ ;
- $S_i$ : è l'area della superficie  $i$ -esima misurato in  $m^2$ .

Questo indice è una misura caratteristica della perdita di calore dell'edificio attraverso l'involucro. Un valore  $U_0$  più grande significa più basse prestazioni termiche dell'involucro edilizio; infatti, un valore  $U_0$  inferiore suggerisce solitamente un elevato livello di isolamento. Il valore  $U_0$  è utile per prevedere il comportamento termico globale di un intero edificio piuttosto che basarsi sulle proprietà di un singolo elemento costruttivo.

### 13.4 Le simulazioni e la validazione

Per ottenere modelli che rappresentano una condizione media relativa ai guadagni solari e per moderare l'aumento della temperatura nello spazio, l'edificio tipo per ogni stato è stato simulato 8 volte variandone l'orientamento di  $45^\circ$  ed eseguendo un totale di 168 simulazioni.

Per validare i modelli simulati, per ogni stato e città sono stati confrontati i dati ottenuti dalle simulazioni con i dati di cui alla Tabella 13.5, fornendo errori, riportati in Tabella 13.12, che non superano il 10% rispetto al valore empirico.

|             | Simulazione TRNSYS<br>Domanda termica<br>[kwh/m <sup>2</sup> ] | Dati empirici<br>Domanda termica<br>[kwh/m <sup>2</sup> ] | Errore relativo<br>[%] |
|-------------|--|---|------------------------|
| Svezia      | 55,3   | 50,8  | 9%                     |
| Germania    | 68   | 66,8  | 2%                     |
| Regno Unito | 59,5   | 62,5  | 5%                     |
| Belgio      | 89,5   | 87,5  | 2%                     |
| Francia     | 61,2   | 65,5  | 7%                     |
| Italia      | 32,9   | 32,5  | 1%                     |
| Spagna      | 12,3   | 12,5  | 2%                     |

Tabella 13.12 - Confronto tra dati simulate e dati empirici.



### 13.5 Modello di regressione $H_d$ -HDD

Attraverso le analisi condotte e gli approfondimenti prima evidenziati, è stato possibile lo sviluppo di sette correlazioni che possono descrivere brevemente il comportamento termico dell'edificio in qualsiasi condizione climatica. I risultati ottenuti sono mostrati in Figura 13.8, che permette di determinare una correlazione generale valida per le città Europa, dove  $H_d$  è una funzione logaritmica dei valori HDD:

$$H_d = 29.329 \ln(\text{HDD}) - 192.65 \quad [96]$$

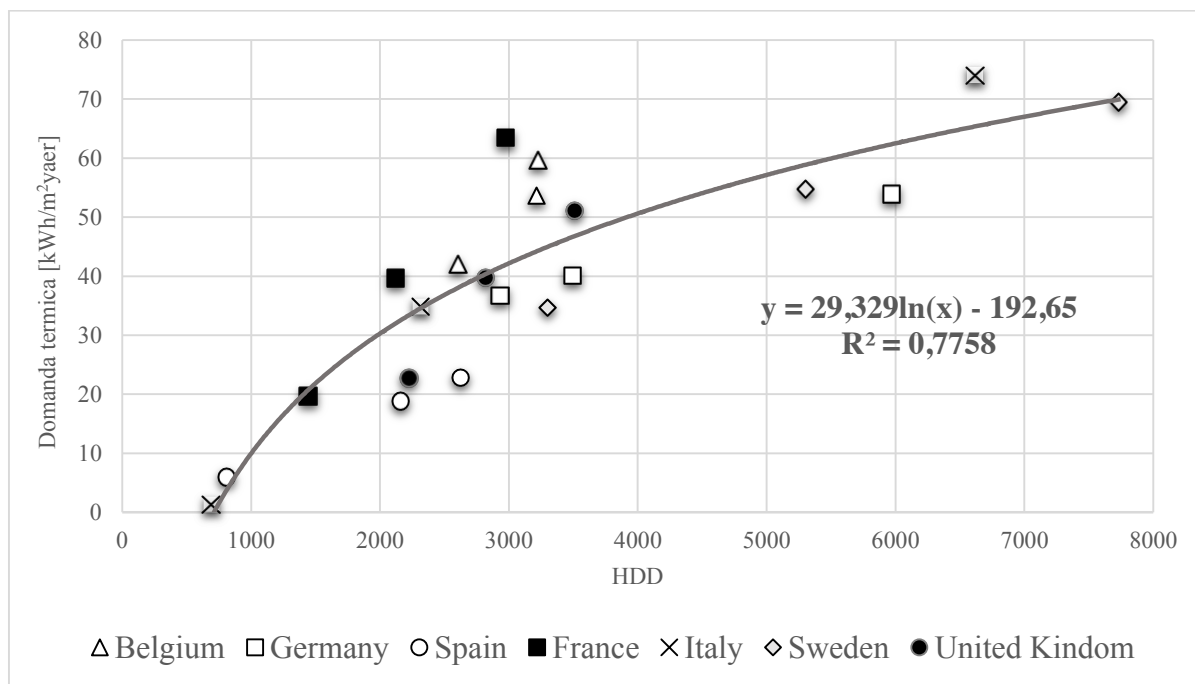


Figura 13.8 - Correlazione tra HDD e domanda termica.

Questa correlazione è caratterizzata da un coefficiente di determinazione ( $R^2$ ) o R-quadro vicino al valore 80%. Essa permette ad un Energy manager di stimare il valore di  $H_d$  con la sola conoscenza del valore di HDD di un edificio per uffici situato nella zona europea con caratteristiche medio-elevate di prestazioni termo-fisiche ( $m^2_{finestra} / m^2_{parete}$ ;  $m^2_{tetto} / m^2_{parete}$ ) e valori di carico interno standard.

Sono state ulteriormente elaborate correlazioni che sono molto più accurate per i singoli paesi, con valori  $R^2$  vicini al 90% (Figure 13.9-13.12); la migliore correlazione è per la Svezia (Figura 13.11), la peggiore risulta per il Belgio (Figura 13.9).

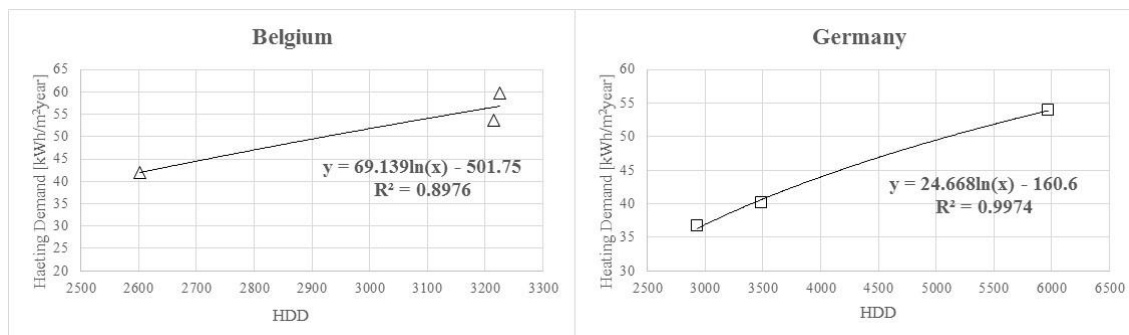


Figura 13.9 - Correlazione tra HDD e domanda termica per Belgio e Germania.

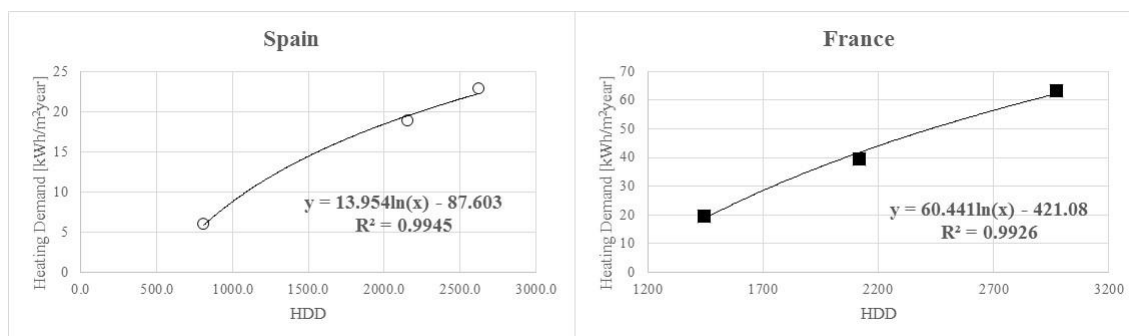


Figura 13.10 - Correlazione tra HDD e domanda termica per Spagna e Francia.

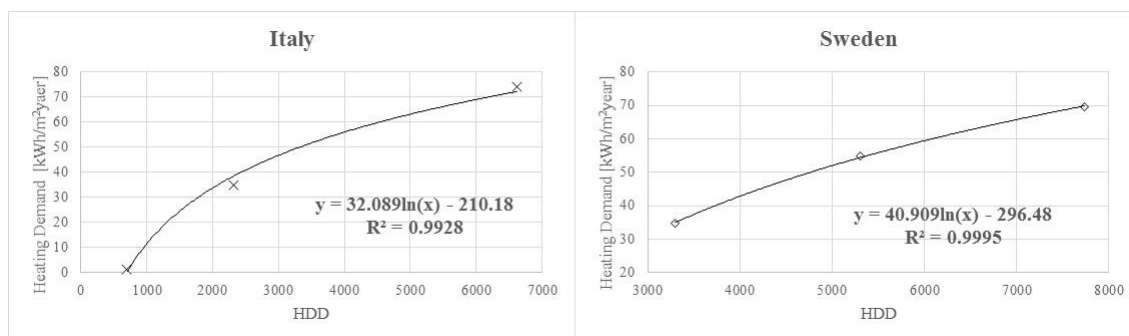


Figura 13.11 - Correlazione tra HDD e domanda termica per Italia e Svezia.

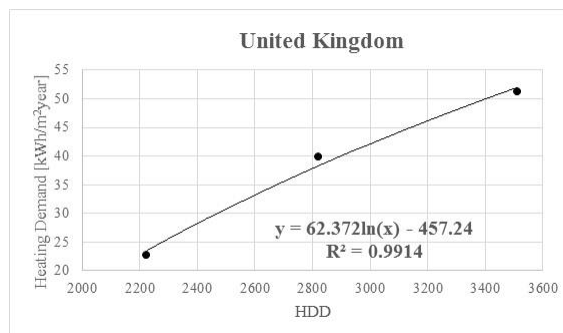


Figura 13.12 - Correlazione tra HDD e domanda termica per Regno Unito.

Inoltre, è stato possibile estrapolare tre correlazioni generali per le macro aree settentrionali, centrali e meridionali dell'Europa (Figure 13.13-13.15).

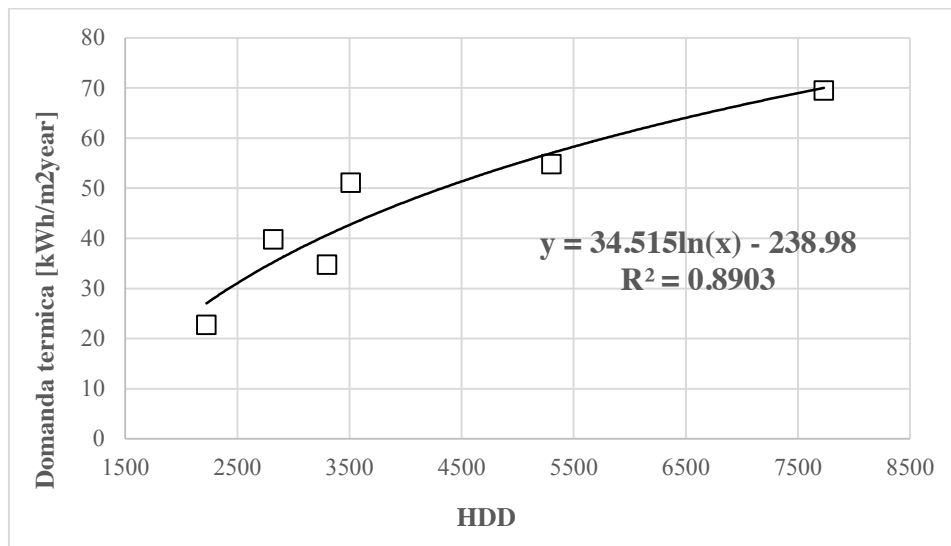


Figura 13.13 - Correlazione tra HDD e domanda termica per il nord Europa.

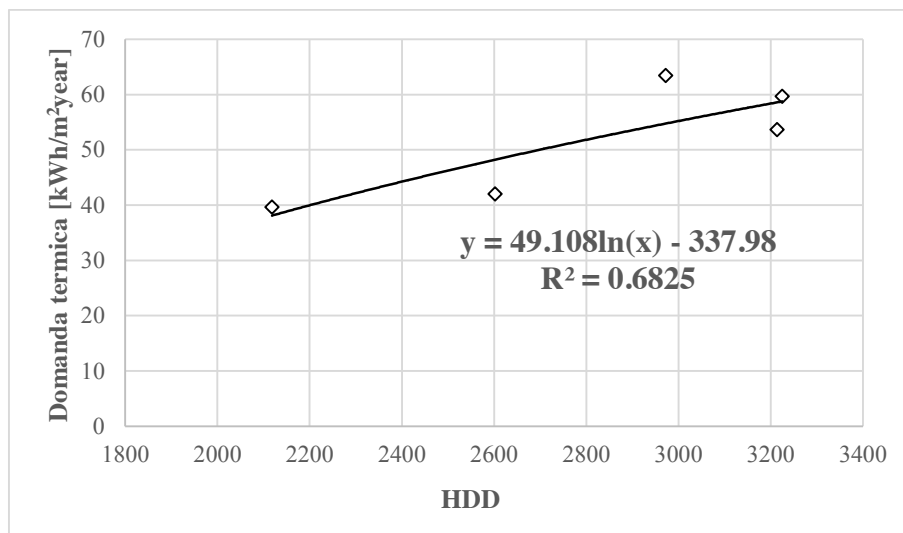


Figura 13.14 - Correlazione tra HDD e domanda termica per il centro Europa.

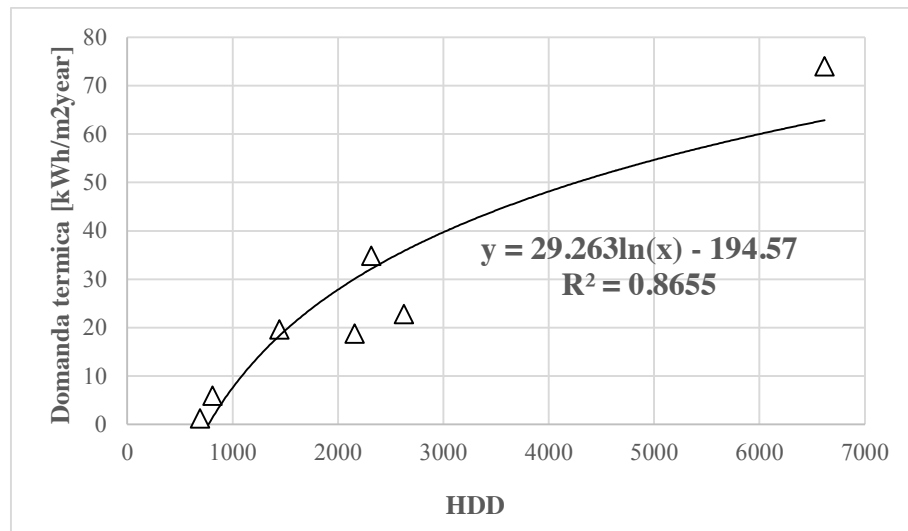


Figura 13.15 - Correlazione tra HDD e domanda termica per il sud Europa.

La prima correlazione illustrato in Figure 13.13 si riferisce alla Svezia ed al Regno Unito, ed è caratterizzata da un  $R^2 = 0,89$ . La seconda correlazione è per l'Europa centrale, tale correlazione è la peggiore delle tre visto il valore di  $R^2 = 0,68$  (Figure 13.14). Per l'Europa meridionale, è stata determinata una buona correlazione con un  $R^2 = 0,86$  (Figure 13.15).

Le correlazioni sviluppate, seppur semplici, forniscono indici di accuratezza elevati e sono il prodotto di applicazioni su modelli dettagliati e dinamici di edifici. Esse potranno quindi consentire ai energy manager o pianificatori energetici di valutare in generale la domanda di energia per uso riscaldamento di un edificio uso uffici in qualsiasi area europea.

## CAPITOLO 14 *Conclusioni*

Il lavoro di ricerca partendo da un'analisi dello stato della ricerca nel campo della predizione di indici e consumi energetici a scala locale, ha inizialmente approfondito le metodologie e i software che si prestano a tali studi. Successivamente, sono stati individuati i *Casi studio* da implementare e sono stati costruiti i DataSet per ognuno di essi, attraverso la ricognizione ed il reperimento di dati ed informazioni necessari per i fini delle predizioni. Una volta organizzato il DataSet sono stati implementati i singoli *Casi studio*, andando a sua volta a generare nove modelli predittivi basati su Neural Network o serie storiche; rappresentativi, a loro volta, dei migliori parametri utilizzati per la definizione della bontà del modello.

Nel complesso i modelli implementati con Serie Storiche hanno mostrato valori di nMAE e nRMSE non superiori al 14% e 17% (Tabella 14.1)

| Caso studio                               | 10.1   |        |        |        |        |       | 11      |         |        | 12     |
|---|--------|--------|--------|--------|--------|-------|---------|---------|--------|--------|
| Numero casi osservati                     | 18     | 18     | 18     | 18     | 18     | 18    | 1360    | 1347    | 136    | 8760   |
| Normalized Mean absolute error (nMAE)     | 1,90 % | 3,43 % | 6,61 % | 3,30 % | 0,77%  | 2,65% | 9,36 %  | 13,73 % | 6,67 % | 6,67 % |
| Normalized Root mean square error (nRMSE) | 2,08 % | 4,17 % | 7,30 % | 4,03 % | 0,95 % | 3,23% | 12,06 % | 17,23 % | 8,83 % | 8,83 % |

Tabella 14.1 – Indici di bontà dei modelli serie storiche per i casi del parafraso 10.1 e capitoli 11 e 12.

Nei modelli regressivi basati su Neural Network, il nMAE non ha superato il valore di 12 % mentre il nRMSE quello del 16,5% per il model fit (Tabella 14.2-14.3). Mentre per la prediction il nMAE non supera il valore di 38,5% ed il nRMSE quello di 53,5% (Tabella 14.4-14.5). Tuttavia questi valori elevati sono basati su un campione dati per la convalida del modello pari a 5 (scala regionale) che certamente ne limita l'effettiva veridicità. Come si è detto nel corso del lavoro, tale anomalia è stata superata con l'utilizzo dei dati disaggregati su base provinciale. Infatti, con un campione di verifica pari ad 11 i valori massimi di nMAE e di nRMSE scendo rispettivamente sotto il 10% ed il 16%.

| Caso studio                               | 10.1   |        |         |        |        |        | 10.2   |        |        |        |
|---|--------|--------|---------|--------|--------|--------|--------|--------|--------|--------|
| Numero casi osservati                     | 18     | 18     | 18      | 18     | 18     | 18     | 43     | 43     | 43     | 43     |
| Normalized Mean absolute error (nMAE)     | 4,34 % | 4,86 % | 11,83 % | 6,02 % | 1,08 % | 4,20 % | 1,07 % | 3,88 % | 2,88 % | 1,18 % |
| Normalized Root mean square error (nRMSE) | 5,69 % | 6,50 % | 16,41%  | 7,14%  | 1,48%  | 5,42%  | 1,44 % | 5,08 % | 3,80 % | 1,65 % |

Tabella 14.2 – Indici di bontà del model fit modelli NN per i casi dei paragrafi 10.1 e 10.2.

| Caso studio                               | 10.3   |        | 11      |        |        | 12     |
|---|--------|--------|---------|--------|--------|--------|
| Numero casi osservati                     | 2464   | 2464   | 1169    | 1169   | 1169   | 28051  |
| Normalized Mean absolute error (nMAE)     | 0,13 % | 0,80 % | 8,91 %  | 7,03 % | 1,53 % | 6,67 % |
| Normalized Root mean square error (nRMSE) | 0,34 % | 1,50 % | 11,74 % | 9,16 % | 6,22 % | 8,83 % |

Tabella 14.3 – Indici di bontà del model fit dei modelli NN per i casi del paragrafo 10.3 e capitoli 11, 12.

| Caso studio                               | 10.1   |        |         |        |        |         | 10.2   |         |        |        |
|---|--------|--------|---------|--------|--------|---------|--------|---------|--------|--------|
| Numero casi osservati                     | 5      | 5      | 5       | 5      | 5      | 5       | 11     | 11      | 11     | 11     |
| Normalized Mean absolute error (nMAE)     | 17,3 % | 9,05 % | 38,30 % | 4,21 % | 7,28 % | 23,47 % | 1,11 % | 8,96 %  | 6,22 % | 2,00 % |
| Normalized Root mean square error (nRMSE) | 22,2 % | 9,36 % | 53,16 % | 5,43 % | 13,3 % | 28,9 %  | 1,36 % | 15,26 % | 9,43 % | 2,40 % |

Tabella 14.4 – Indici di bontà della prediction dei modelli NN per i casi dei paragrafi 10.1 e 10.2.

| Caso studio                               | 10.3   |        | 11      |         |         | 12     |
|---|--------|--------|---------|---------|---------|--------|
| Numero casi osservati                     | 616    | 616    | 292     | 292     | 292     | 28051  |
| Normalized Mean absolute error (nMAE)     | 0,13 % | 0,78 % | 9,30 %  | 7,92 %  | 15,94 % | 6,69 % |
| Normalized Root mean square error (nRMSE) | 0,30 % | 1,46 % | 12,16 % | 11,21 % | 23,21 % | 8,82 % |

Tabella 14.5 – Indici di bontà della prediction dei modelli NN per i casi del paragrafo 10.3 e capitoli 11 e 12.

Nel modello basato su regressione logaritmica, il cui obiettivo è stato quello di sviluppare correlazioni semplici per valutare la domanda di energia per uso riscaldamento di un edificio uso uffici in qualsiasi area del continente europeo, i valori dell'indice  $R^2$  per singolo paese non sono stati inferiori al valore 0,89. Mentre per gruppi di Stati, appartenenti a zone climatiche simili, l'indice  $R^2$  ha fatto registrare il valore 0,68 come valore peggiore.

Nello specifico, gli studi sulla predizione dei consumi energetici, a sua volta suddivisi in tre sotto casi, hanno confermato la migliore resa dei sistemi basati su Neural Network quando si ha a disposizione un DataSet ben strutturato e rappresentativo di un campione di dati numeroso. Le analisi su DataSet provinciali sono quindi risultati più vantaggiose, rispetto a quelle con DataSet regionale. Infatti, grazie ai dati disaggregati, nel caso della Sicilia in nove sotto insiemi (province), sono stati utilizzati come input molti più dati per l'allenamento della rete e di conseguenza sono migliorati i parametri di bontà del modello predittivo. È quindi preferibile, anche nel caso di predizione degli indici energetici regionali, procedere attraverso la predizione dei parametri provinciali, e successivamente per aggregazione addivenire al dato Regionale. Per quanto riguarda invece i dati Comunali, i risultati ottenuti hanno evidenziato come la buona organizzazione del DataSet garantisca al contempo buoni parametri di bontà della rete (nMAE e nRMSE). È quindi evidente come i modelli su serie storiche siano consigliabili esclusivamente per analisi relative ad un unico soggetto (Regione), e sconsigliabile nel caso di più province e comuni. Questo poiché l'organizzazione del DataSet genera una perdita

della sequenza storica che non rende applicabile alcun modello di previsione basato esclusivamente sulla periodicità di eventi e stagionalità dell'output.

A livello regionale e provinciale è evidente che per il modello predittivo Neural Network la scelta di dati di input, quali popolazione, valore aggiunto e temperatura: lo caratterizzano ottimamente per la previsione dei consumi nel settore industriale; in maniera buona per la previsione nei settori residenziale, terziario, servizi non vendibili ed attività produttive; non risultano prestazioni ottimali per la quantificazione dei consumi nel settore agricolo.

Nel modello di previsione NN su scala comunale, gli indici utilizzati come input al modello superficie territoriale, altitudine, temperatura cumulata, gradi giorno, radiazione solare, popolazione, lunghezza rete gas, indice di comune montano; risultano fornire buoni livelli di affidabilità e caratterizzazione, più per la previsione del consumo di energia elettrica che del consumo di gas naturale.

Nel caso studio per la previsione della produzione di energia elettrica da FER si è potuto evidenziare che dal confronto dei due modelli, il modello basato su Neural Network rispetto al modello serie storiche ha reso migliori indici di bontà per i target relativi alla produzione da Fotovoltaico ed Eolico.

Analizzando invece i due modelli implementati per la previsione del prezzo zonale è evidente come il modello serie storiche fornisca, seppur di poco, migliori indici rispetto a quello basato su modello Neural Network. La quasi equivalenza dei due modelli in termini di affidabile previsione, è essenzialmente dovuta dalla buona organizzazione del DataSet, (serie continue e pluriennali) e dall'elevata inerzia di un sistema energetico su scala regionale che generalmente presenta caratteristiche di stagionalità e ripetitività. Tali caratteristiche rendono la previsione con modelli di serie storiche molto accurate e con bassi errori, facendo sì che tali modelli siano competitivi rispetto a quelli Neural Network.

A conclusione del lavoro, possiamo affermare che per la previsione di indicatori energetici relativi a sistemi energetici su scala regionale, provinciale e locale, i modelli con serie storiche risultano superiori ai modelli Neural Network nel caso il DataSet sia costituito da un numero basso di dati. In tal senso la previsione assume più una veste di previsione del trend che del dato effettivo. Nel caso di DataSet ben organizzato i risultati per modelli Neural Network risultano essere più affidabili e precisi, fornendo indici di bontà quanto più elevati al crescere del numero di campioni del DataSet.

Si deve anche evidenziare che al crescere del campione nel DataSet, quando questo non perde le caratteristiche di serie storica, gli indici di bontà del modello Neural Network e del modello serie storiche si eguagliano a causa di fattori di periodicità e stagionalità che caratterizzano i dati energetici di macro aree come comuni, province e regioni.

Infine, è bene evidenziare come gli indicatori di input utilizzati nelle analisi condotte e relativi a condizioni sociali, economiche, geomorfologiche, climatiche ed impiantistiche rendono i modelli, predittivi sviluppati con approccio regressivo (Neural Network e regressione logaritmica) appropriati,

nella maggior parte dei casi, alla predizione dei consumi energetici e della produzione da fonti energetiche rinnovabili su scala locale, provinciale e regionale.



## ***Bibliografia***

- [1] Pacchetto “unione dell’energia” Bruxelles, 25.2.2015 com(2015) 80 final Commissione europea;
- [2] L’Unione dell’Energia: energia sicura, sostenibile, competitiva e a prezzi accessibili per tutti gli europei - Bruxelles, 25.2.2015 - Commissione Europea.
- [3] Relazione della Commissione al Parlamento europeo e al Consiglio - Progressi nella realizzazione degli obiettivi di kyoto e di europa 2020 (a norma dell’articolo 21 del regolamento (UE) n. 525/2013 del Parlamento europeo e del Consiglio, del 21 maggio 2013, relativo a un meccanismo di monitoraggio e comunicazione delle emissioni di gas a effetto serra e di comunicazione di altre informazioni in materia di cambiamenti climatici a livello nazionale e dell’Unione europea e che abroga la decisione n. 280/2004/CE) /\* COM/2013/0698 final.
- [4] Il quadro per le politiche dell’energia e del clima all’orizzonte 2030 Bruxelles, 24.10.2014 - Commissione europea.
- [5] E. Moreci, G. Ciulla and V. Lo Brano. “The Energy System of Sicilian Region, Italy.” ICRERA 2015: Proceedings of the 4th International Conference on Renewable Energy Research and Applications. ICRERA, 2015.
- [6] Kadoshin, Shiro, Takashi Nishiyama, and Toshihide Ito. “The trend in current and near future energy consumption from a statistical perspective.” Applied Energy 67.4 (2000): 407-417.
- [7] Ozturk, Harun Kemal, et al. “Electricity estimation using genetic algorithm approach: a case study of Turkey.” Energy 30.7 (2005): 1003-1012.
- [8] [Tso, Geoffrey KF, and Kelvin KW Yau. “Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks.” Energy 32.9 (2007): 1761-1768.
- [9] Beccali, M., et al. “Short-term prediction of household electricity consumption: Assessing weather sensitivity in a Mediterranean area.” Renewable and Sustainable Energy Reviews 13.8 (2008): 2040-2065.
- [10] Azadeh, A., and S. Tarverdian. “Integration of genetic algorithm, computer simulation and design of experiments for forecasting electrical energy consumption.” Energy Policy 35.10 (2007): 5229-5241.
- [11] Azadeh, Ali, et al. “Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption.” Applied Mathematics and Computation 186.2 (2007): 1731-1741
- [12] Mayer, Audrey L. “Strengths and weaknesses of common sustainability indices for multidimensional systems.” Environment international 34.2 (2008): 277-291.

- [13] A. Azadeh, S. F. Ghaderi, and S. Sohrabkhani. “Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors.” *Energy Conversion and Management* 49.8 (2008): 2272-2278.
- [14] Shiwei Yu, Ke Wang,, Yi-Ming Wei “ A hybrid self-adaptive Particle Swarm Optimization–Genetic Algorithm–Radial Basis Function model for annual electricity demand prediction.” *Energy Conversion and Management* 91 (2015): 176-185.
- [15] Senatro Di Leo, Carmelina Cosmi, Maria Ragosta “ An application of multivariate statistical techniques to partial equilibrium models outputs: The analysis of the NEEDS-TIMES Pan European model results”. *Renewable and Sustainable Energy Reviews* 49 (2015): 108-120.
- [16] Concepts and methods in energy statistics, with special reference to energy accounts and balances: a technical report, *Studies in Methods, Series F, No. 29*, United Nations, New York, 1982;
- [17] Energy statistics: definitions, units of measure and conversion factors, *Studies in Methods, Series F, No. 44*, United Nations, New York, 1987.
- [18] Energy statistics: a manual for developing countries, *Studies in Methods, Series F, No. 56*, United Nations, New York, 1991.
- [19] Energy Statistics Manual, OECD/IEA/EUROSTAT, Paris, 2004.
- [20] Direttiva 2001/77/CE del Parlamento europeo e del Consiglio, del 27 settembre 2001, sulla promozione dell'energia elettrica prodotta da fonti energetiche rinnovabili nel mercato interno dell'elettricità.
- [21] Direttiva 2004/8/CE del Parlamento europeo e del Consiglio, dell'11 febbraio 2004, sulla promozione della cogenerazione basata su una domanda di calore utile nel mercato interno dell'energia.
- [22] Direttiva 2002/91/CE del Parlamento europeo e del Consiglio, del 16 dicembre 2002, sul rendimento energetico nell'edilizia.
- [23] Direttiva 2006/32/CE del Parlamento europeo e del Consiglio, del 5 aprile 2006, concernente l'efficienza degli usi finali dell'energia e i servizi energetici.
- [24] Direttiva 2005/32/CE del Parlamento europeo e del Consiglio, del 6 luglio 2005, relativa all'istituzione di un quadro per l'elaborazione di specifiche per la progettazione ecocompatibile dei prodotti che consumano energia; impongono agli Stati membri di trasmettere dati quantitativi sul consumo energetico. Rilevano la necessità di una migliore interfaccia tra i dati sull'energia e le indagini statistiche pertinenti, come il censimento della popolazione e degli alloggi e i dati sui trasporti.
- [25] Regolamento (ce) n. 1099/2008 del Parlamento europeo e del Consiglio del 22 ottobre 2008 relativo alle statistiche dell'energia.
- [26] D.M. del 14 gennaio 2012 del Ministero dello Sviluppo economico, che ha approvato la metodologia che, nell'ambito del sistema statistico nazionale in materia di energia, è applicata per rilevare i dati necessari a

misurare il grado di raggiungimento degli obiettivi nazionali in materia di quote dei consumi finali lordi di elettricità, energia per il riscaldamento e il raffreddamento, e per i trasporti coperti da fonti energetiche rinnovabili.

- [27] D.M. del 11 maggio 2015 del Ministero dello Sviluppo economico, di approvazione della metodologia che, nell'ambito del sistema statistico nazionale, è applicata per rilevare i dati necessari a misurare il grado di raggiungimento degli obiettivi regionali, in attuazione dell'articolo 40, comma 5, del decreto legislativo 3 marzo 2011, n. 28.
- [28] Ciulla, G., V. Lo Brano, and E. Moreci. "Degree days and building energy demand." 3rd Southern African Solar Energy Conference, South Africa, 11-13 May, 2015., 2015.
- [29] CENTRA, MASSIMO. "La previsione dei consumi elettrici: un'applicazione al settore del trasporto ferroviario." (2009).
- [30] Anastasakis, L., and N. Mort. "The development of self-organization techniques in modelling: a review of the group method of data handling (GMDH)." (2001).
- [31] A.G. Ivakhnenko, G.I. Krotov and V.S. Stepashko, "Harmonic and exponential harmonic GMDH algorithms. Part 2. Multilayer algorithms with and without calculation of remainders", Soviet Automatic Control c/c of Avtomatika, vol.16, no.1, pp.1-9, 1983.
- [32] A.G. Ivakhnenko, G. Petrache and M.S. Krasys'tkyy, "A GMDH algorithm with random selection of pairs", Soviet Automatic Control c/c of Avtomatika, vol.5, no.4, pp.23-30, 1972.
- [33] Valerio Lo Brano, Giuseppina Ciulla, and Mariavittoria Di Falco. "Artificial neural networks to predict the power output of a PV panel." International Journal of Photoenergy 2014 (2014).
- [34] Note metodologiche Volume A - S.I.E.R. Sistema Informativo Energetico Regionale Regione Sicilia, Bilanci Energetici Regionali dal 1988 al 1999.
- [35] Manuale dell'utente di IBM SPSS Statistics 21 Core System.
- [36] Ministero dello Sviluppo Economico – Statistiche dell'Energia – "Dati dei consumi provinciali di gas naturale." " per gli anni: 2010, 2011, 2012 and 2013.
- [37] Ministero dello Sviluppo Economico – Statistiche dell'Energia – "Vendite provinciali di benzine, gasolio, olio combustibile, GPL e lubrificanti." per gli anni: 2010, 2011 and 2012.
- [38] Terna S.p.a., "Dati statistici sull'energia elettrica in Italia" per gli anni dal 1997 al 2014.
- [39] Gestore dei Servizi Energetici – GSE S.p.A., "Evoluzione del Conto Energia".
- [40] Gestore dei Servizi Energetici – GSE S.p.A., "Rapporto statistico impianti a fonti rinnovabili" per gli anni dal 2008 al 2012.

- [41] Gestore dei Servizi Energetici – GSE S.p.A., “Bollettino informativo sull’energia da fonti rinnovabili” per: 2° semestre 2008, 2° semestre 2009, 2° semestre 2010, 2° semestre 2011, 2° semestre 2012, 1° semestre 2013.
- [42] Gestore dei Servizi Energetici – GSE S.p.A., “Dati principali relativi alla potenza, alla numerosità ed alla gestione commerciale degli impianti fotovoltaici” per gli anni dal 2006 al 2013 e per singolo mese.
- [43] Italian Greenhouse Gas Inventory 1990-2011, Istituto superiore per la protezione e la ricerca ambientale (ISPRA).
- [44] Censimento della popolazione e delle abitazioni – ISTAT.
- [45] Tav. 2 - Popolazione attiva per settore di attività economica e valore aggiunto complessivo e pro capite per provincia - Fondazione “Istituto Guglielmo Tagliacarne” – Unioncamere.
- [46] Osservatorio Regionale e Ufficio Statistico per l’Energia – Regione Siciliana.
- [47] Esiti dei mercati e statistiche - Gestore mercati energetici (GME).
- [48] Commission of the European Communities (1992). Proposal for a Council Directive Introducing a Tax on Carbon Dioxide Emissions and Energy.COM (92) 226 Final. Commission of the European Communities (2008), 20 20 By 2020 — Europe’s.
- [49] Climate Change Opportunity, Brussels, COM (2008) 30 Final. Department of Energy and Climate Change, 2011. Planning our Electric Future, a White Paper for Secure, Affordable and Low-carbon Electricity. The Stationery Office, London, CM 8099.
- [50] European Parliament, 1996.Directive 96/92/EC of the European Parliament and of the Council of 19 December 1996 Concerning Common Rules for the Internal Market in Electricity. Official Journal of the European Union L 027, 20–29.
- [51] European Parliament, 1998. Directive 98/30/EC of the European Parliament and of the Council of 22 June 1998 Concerning Common Rules for the Internal Market in Natural Gas. Official Journal of the European Union, L 204, 1–12.
- [52] European Parliament, 2003a. Directive 2003/87/EC as amended by 2009/29/EC reviewing the EU emission trading scheme by defining an emissions cap and harmonising allocation of allowances to companies. Official Journal of the European Union, L 140, 63–87.
- [53] European Parliament, 2009a. Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing directives 2001/ 77/EC and 2003/30/EC. Official Journal of the European Union 140, 16–62L 140, 16–62.
- [54] European Commission (2013), Green Paper: A 2030 Framework for Climate and Energy Policies. COM (2013) 169 Final.

- [55] Beccali, M., Bonomolo, M., Ciulla, G., Galatioto, A., & Brano, V. L. (2015). Improvement of energy efficiency and quality of street lighting in South Italy as an action of Sustainable Energy Action Plans. The case study of Comiso (RG).Energy.
- [56] Eskin, Nurdil, and Hamdi Türkmen. “Analysis of annual heating and cooling energy requirements for office buildings in different climates in Turkey.” *Energy and Buildings* 40.5 (2008): 763-773.
- [57] Christenson, M., Manz, H., & Gyalistras, D. (2006). Climate warming impact on degree-days and building energy demand in Switzerland. *Energy Conversion and Management*, 47(6), 671-686.
- [58] Lee, K., Baek, H. J., & Cho, C. (2014). The estimation of base temperature for heating and cooling degree-days for South Korea. *Journal of Applied Meteorology and Climatology*, 53(2), 300-309.
- [59] Kim, J. H., & Suh, S. J. (2006). The Demand Expectation of Heating & Cooling Energy in Buildings According to Climate Warming. *Journal of the Korean Solar Energy Society*, 26(3), 119-125.
- [60] Yi-Ling, H., Hai-Zhen, M., Guang-Tao, D., & Jun, S. (2014). Influences of urban temperature on the electricity consumption of Shanghai. *Advances in Climate Change Research*, 5 (2), 74-80.
- [61] TRNSYS: Transient System Simulation Tool, <http://www.trnsys.com/>.
- [62] Edelgard Gruber, Stefan Plessner, Rinto Dusée, Ilias Sofronis, Pedro Lima, Philippe Rivière, Anne Rialhe, “Deliverable D 26: Report on the Project Results - EL-TERTIARY Monitoring Electricity Consumption in the Tertiary Sector EIE/05/144/SI2.419658 Intelligent Energy – Europe (IEE)” – (2008).
- [63] Eicker, Ursula. “Energy Efficient Buildings with Solar and Geothermal Resources” - Wiley (2014).
- [64] Eurostat, the statistical office of the European Union - 2920 Luxembourg.
- [65] <http://nesa1.uni-siegen.de/wwwextern/idea/keytopic/3.htm> (April 22nd, 2015).
- [66] European Insulation Manufacturers Association – EURIMA.
- [67] Meteonorm - Global Meteorological Database - Version 7 - Software and Data for Engineers, Planners and Education.

